Global Happiness Analysis

Arvind Kandala, Ammy Lin, Lingyue Hao, Shelly Cao

Data Science Memo

The primary stakeholder for this memo is the Administrator of the United Nations Development Program (UNDP), who oversees global development strategy and the prioritization of international aid packages. This stakeholder is well-versed in core metrics like those in the World Happiness Report, but does not work directly with complex statistical interaction modeling. As such, they require clear strategic interpretations rather than technical detail. They care about this analysis because understanding whether the drivers of happiness are strictly dependent on economic and political prerequisites can help determine whether aid resources should be shifted toward foundational stabilization before funding broader health and social wellness initiatives.

Executive Summary

Global policy institutions increasingly prioritize population well-being, yet the structural factors that drive national happiness remain only partly understood. This analysis draws from The World Happiness Report, which aggregates responses from the Gallup World Poll to measure subjective well-being. These responses are based on the Cantril ladder question, which asks respondents to think of a ladder, where the best possible life for them is a 10, and the worst possible life is a 0, and rate their current lives on that scale. While the report identifies GDP per capita and healthy life expectancy (both measured as 0 to 1 indices rather than raw dollars or years), freedom to make life choices, generosity, and social support as major contributors to global differences in life satisfaction, the ways these factors interact are still not fully resolved (World Happiness Report, 2024). Similarly, cross-country analyses compiled by Our World in Data show that both economic and health conditions are strongly associated with national well-being, underscoring the need for models that evaluate their combined effects (Roser &

Ortiz-Ospina, 2024). To investigate these relationships, this analysis employs two statistical approaches: first, a multiple linear regression testing whether the relationship between life expectancy and national happiness depends on GDP per capita; and second, an ordinal logistic regression assessing whether a country's levels of generosity and freedom are associated with an increased probability of belonging to a higher regional happiness group. More specifically, generosity is defined as the national average share of people who answered yes to having donated money to a charity in the past month (Helliwell et al., 2025).

Our analysis reveals that the drivers of happiness depend heavily on structural context. We find that economic stability acts as a prerequisite for the benefits of public health: extending life expectancy significantly boosts national happiness only in nations with high GDP per capita, whereas in lower-income settings, longer lives do not translate into higher reported well-being. Similarly, we find that social virtues like generosity only correlate with higher regional happiness when citizens possess high levels of personal freedom. Consequently, we recommend that you view economic development and civil liberties not just as independent goals, but as the necessary foundations required to enjoy the well-being benefits of healthcare improvements and social capital.

Decisions To Be Made

As public policy advisers, you face a critical decision regarding resource allocation: should you prioritize policy interventions focused on public health infrastructure, or continue prioritizing pure economic expansion? Specifically, you must determine whether to shift resources toward initiatives that extend life expectancy, given our analysis of its association with national happiness after adjusting for GDP per capita. Additionally, you must consider how your promotion of civil liberties and charitable incentives impacts happiness.

When reviewing our findings, we suggest you adopt a "hierarchy of needs" approach to policy. Since the data indicate that longevity yields diminishing returns on happiness in low-GDP contexts, your immediate funding in developing regions should perhaps prioritize economic stabilization and infrastructure over purely life-extending medical interventions. Similarly, given that social incentives (generosity) appear ineffective without personal autonomy, you may need to enact political reforms protecting civil liberties before attempting to build social capital. Conversely, in wealthier and freer nations with a strong political and economic foundation, investing in preventive care and pro-social community programs would offer you a much higher return on societal well-being than in developing countries.

Key Findings

The regression results indicate that several factors are meaningfully associated with national happiness, with GDP per capita, family support, and freedom showing the most robust positive coefficients in the full model. Although the main effect of life expectancy alone cannot be interpreted in isolation due to the presence of interaction terms, the significant positive interaction between life expectancy and GDP suggests that the relationship between longevity and happiness depends on a country's economic conditions. In particular, increases in life expectancy correspond to larger gains in happiness in higher-GDP countries, whereas the association is substantially weaker in lower-GDP contexts.

For regional happiness, the results again indicate that the interaction between generosity and freedom—not the individual predictors in isolation—is the key driver of the relationship. Because the model includes an interaction, the coefficients for generosity and freedom represent their associations only when the other variable equals zero, a value that is not meaningful in this context. Thus, the main effects cannot be interpreted as the direct influence of generosity or freedom on regional happiness. What the model does show is that the combined effect of generosity and freedom is highly significant: generosity is associated with higher regional happiness when freedom is relatively high, but it is weakly related—or even negatively related—when freedom is low. In other words, generosity contributes more strongly to regional well-being in environments where people have sufficient autonomy to act on prosocial tendencies, reinforcing that the interaction, rather than the individual predictors, explains the observed pattern.

Interpretation and Implications

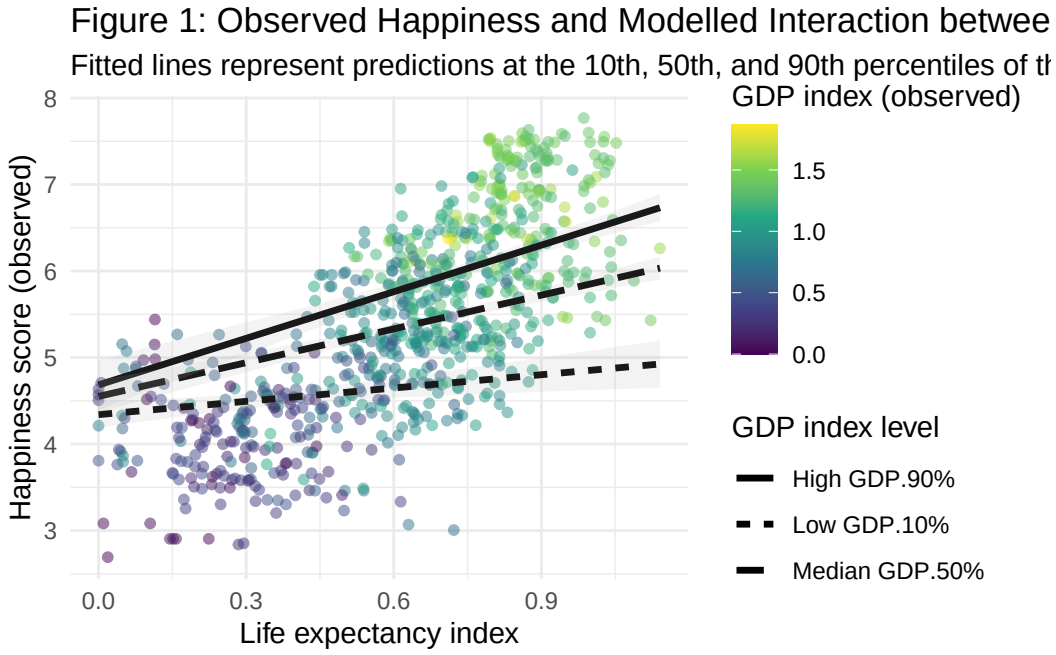


Figure 1 above presents the observed relationship between the life expectancy index and national happiness, together with fitted regression lines that illustrate the estimated interaction between life expectancy and GDP. The horizontal axis displays the life expectancy component from the World Happiness Report. This variable is not measured in years; rather, it is a 0 to approximately 1.03 scaled index representing how much a country's life expectancy exceeds the report's dystopia benchmark—a hypothetical minimum reference level used across all countries.

The vertical axis reports the observed happiness score, which ranges from 0 to 10. Each point corresponds to a country–year observation, and its colour shows the observed value of the GDP component index. Similar to life expectancy, the GDP index is not GDP per capita in dollars. Instead, it is a 0 to approximately 1.7 scaled measure indicating how far a country's GDP per capita lies above the shared dystopia baseline. The World Happiness methodology transforms GDP per capita into this index to make all component scores comparable on a common scale.

Superimposed on the scatter plot are three fitted regression lines representing predicted happiness at the 10th, 50th, and 90th percentiles of the GDP index, while all other covariates (family, freedom, trust, generosity, and year) are held at their mean levels. These lines summarize the interaction estimated in the regression model. The fitted line at the 90th percentile of the GDP index has the steepest positive slope, indicating that increases in the life expectancy index are

associated with larger increases in predicted happiness when GDP is relatively high. In contrast, the fitted line at the 10th percentile is noticeably flatter, showing that improvements in the life expectancy index correspond to smaller predicted increases in happiness when GDP levels are low.

For your strategic planning, the key implication is that economic stability acts as a prerequisite multiplier for public health investments. The impact of longevity is not consistent across borders; it relies heavily on the surrounding economic environment. If you focus policy interventions solely on extending life expectancy in developing regions, you may struggle to improve national happiness scores unless you pair them with initiatives that strengthen the economic safety net and social support systems.

Recommendations

With these results in mind, we recommend you pursue policies that enhance the conditions under which increases in life expectancy actually translate into gains in happiness. Specifically:

1. Pair health improvements with more general economic support, particularly in poorer settings. Our findings indicate that longevity contributes little to happiness if economic conditions remain limited.
2. Pair longevity-focused initiatives with broader economic development efforts, as gains in life expectancy translate into higher well-being, primarily when economic conditions are stronger.
3. Account for shifting long-term trends in how income and longevity relate to national well-being, as the influence of these factors has changed over time.
4. Strengthen civil and personal freedoms in low-freedom regions before expanding prosocial or community-based initiatives. The model indicates that generosity has a much stronger positive association with happiness in contexts where individuals have greater autonomy to act on generous impulses.

Conclusion

Global policy institutions often treat the drivers of happiness in isolation, yet our analysis demonstrates that these factors operate within a rigid structural hierarchy. We found that "higher-order" well-being investments, such as extending life expectancy and incentivizing generosity, only yield significant happiness returns in nations that have already secured the

foundational prerequisites of economic stability and personal freedom. Consequently, we advise that you adopt a "hierarchy of needs" decision framework: prioritizing economic and civil rights infrastructure in developing regions, while shifting focus to preventive health and social capital only in wealthier, freer nations. This tiered approach is critical to ensuring that limited resources are not wasted on interventions that the current structural environment cannot yet support.

Technical Appendix

Data

The World Happiness Report is a benchmark that allows us to understand the self-reported well-being of residents of many nations. Collected by Gallup World Poll in the years 2015 through 2019, the poll surveyed professionals in domains including economics, psychology, survey analysis, national statistics, health, and public policy across many countries to help assess and shape policy decisions going forward. The surveyed professionals reported their happiness score and the degree to which GDP per capita, family, life expectancy, freedom, lack of corruption, and generosity contribute to one's happiness in a given year. Each observation represents the average of the self-reported scores of a country's professionals within a given year. The happiness score and all of the predictors are normalized on a scale of 1 through 10. A score of 0 represents the worst possible life or the least degree of importance, and a score of 10 represents the greatest (Kaggle, 2019).

The raw data included five CSV files, one for each year from 2015 through 2019, that needed to be combined. To address this, a Year column was first added to each dataset to distinguish the observations. The original Kaggle files used slightly different column names in different years, even when they represented the same underlying quantity. For example, the 2015–2017 files contain variables such as Happiness Score and Economy (GDP per Capita), whereas the 2018–2019 files use names like Score and GDP per capita. To properly combine the files into one dataset, a mapping of year-specific names to standardized names was established. For instance, Score was renamed to Happiness Score so that all files share the same set of column names in the final dataset.

Another issue encountered was that the 2018 and 2019 files contained columns summarizing uncertainty in the happiness scores, which lacked uniformity and could not be interpreted directly. These were dropped because they were not relevant to the research question at hand, simplifying the model.

Model Analyses

Research Question 1

Is there a significant association between life expectancy and national happiness, after adjusting for GDP per capita, after controlling for family support, freedom, trust in government, generosity, and year?

Model Description

We estimate a multiple linear regression model to study how life expectancy and economic development jointly relate to national happiness. The outcome variable is the country-level happiness score reported in the World Happiness Report, which is constructed relative to a "Dystopia" baseline that represents the lowest observed levels of each well-being component. All component indices, including life expectancy, GDP per capita, family support, freedom, trust, and generosity, take values between 0 and 1 and quantify how far each country lies above this Dystopia reference point. Because the outcome is continuous and the predictors are measured on interval-like scales, a linear regression framework is appropriate. The model relies on the standard assumptions that (1) the conditional relationship between predictors and the outcome is linear, (2) residuals have constant variance across fitted values, (3) residuals are independent across observational units, and (4) residuals are approximately normally distributed.

The primary objective is to examine whether the association between life expectancy and happiness differs across levels of economic development. To address this, the model includes the life expectancy and GDP per capita indices as continuous predictors along with their interaction. In this specification, the individual coefficients on life expectancy and GDP per capita cannot be interpreted as marginal effects; instead, they contribute to a combined effect that depends jointly on both indicators. The interaction term captures whether improvements in life expectancy translate into larger or smaller increases in happiness, depending on a country's economic conditions.

All remaining available well-being components—family support, freedom, trust, and generosity—are included as covariates to reduce confounding. These factors represent social, institutional, and relational environments that plausibly influence both life expectancy and subjective well-being, and adjusting for them allows the model to better isolate the association of interest. Because the dataset spans multiple time points, the model also includes the survey year as a continuous predictor, along with interactions between year and both life expectancy and GDP per capita.

Since countries contribute repeated observations to the dataset, standard errors are computed using country-level cluster-robust variance estimators. This approach addresses the within-country dependence that would otherwise bias classical standard errors and ensures valid inference even in the presence of serial correlation or heteroskedasticity.

Model Results

Table 1: Linear regression model for national happiness (cluster–robust SEs by country)

Variable	Estimate	Std Error	t-value	p-value	2.5% CI	97.5% CI
Intercept	138.048	85.540	1.614	0.107	-29.872	305.967
Life expectancy index	-341.906	168.642	-2.027	0.043	-672.959	-10.853
GDP per capita index	324.225	103.867	3.122	0.002	120.328	528.122
Year	-0.067	0.042	-1.583	0.114	-0.150	0.016
Family index	0.890	0.118	7.542	<0.001	0.658	1.122
Freedom index	1.480	0.274	5.394	<0.001	0.942	2.019
Trust index	0.367	0.560	0.656	0.512	-0.732	1.467
Generosity index	0.350	0.332	1.054	0.292	-0.302	1.001
Life expectancy × GDP per capita	1.195	0.486	2.461	0.014	0.242	2.148
Life expectancy × Year	0.170	0.084	2.028	0.043	0.005	0.334
GDP per capita × Year	-0.161	0.052	-3.116	0.002	-0.262	-0.059
R-squared	0.781					
Adjusted R-squared	0.778					

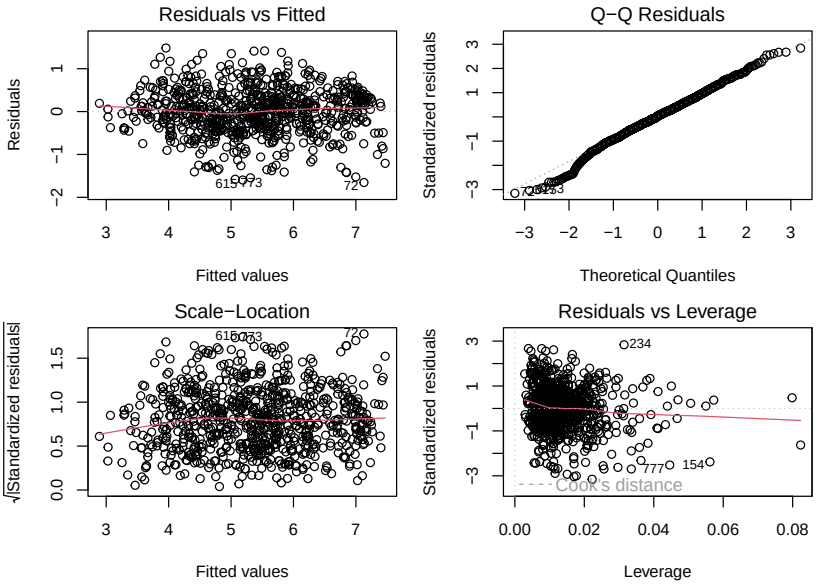
The main effect coefficients in Table 1, such as −341.906 for life expectancy and 324.225 for GDP per capita, appear numerically large because, in a model that includes interaction terms, each main effect represents the estimated association when all interacting variables are equal to 0. In this dataset, such zero values do not occur. The life expectancy and GDP indices are constructed relative to a dystopia reference point and never take the value 0, and the year variable ranges from 2015 to 2019 rather than approaching 0. As a result, the main effect coefficients reflect values at hypothetical predictor combinations outside the empirical data range. They are therefore not meaningful in isolation, and interpretation must instead focus on the interaction terms that determine the estimated relationships within the actual values observed in the dataset.

The interaction terms provide the relevant information for understanding how the association between life expectancy and happiness depends on economic level and time period. The interaction between life expectancy and GDP per capita has an estimated coefficient of 1.195 with a p-value of 0.014 and a confidence interval from 0.242 to 2.148, indicating that the estimated association between life expectancy and happiness becomes stronger at higher GDP levels.

The interaction between life expectancy and year shows a similar pattern. Its estimated coefficient is 0.170 with a p-value of 0.043 and a confidence interval from 0.005 to 0.334. Within the observed period from 2015 to 2019, the model therefore estimates an increasing association between life expectancy and happiness across successive years. These two interactions confirm that the effect of life expectancy cannot be summarized by a single slope and varies across both GDP per capita and year.

Additional covariates also contribute to explaining variation in happiness. The family index and freedom index have statistically significant positive coefficients of 0.890 and 1.480, respectively, indicating that they account for meaningful differences in happiness scores after conditioning on all other predictors. Trust and generosity are not statistically significant in this specification, as their confidence intervals include 0. Overall model fit is strong, with an R-squared of 0.781 and an adjusted R-squared of 0.778, indicating that the model explains a substantial portion of the variation in national happiness.

Model Assessment:



Model diagnostics indicate that standard regression assumptions are reasonably supported, with some mild caveats. The residuals-versus-fitted plot displays residuals scattered without a strong curved trend, indicating no obvious departure from linearity. While the Q–Q plot shows that residuals follow the theoretical line through the central distribution, noticeable deviations in both tails suggest the data departs from perfect normality at the extremes. Similarly, the scale–location plot reveals a slight increase in variability toward the higher end of fitted values, pointing to mild heteroskedasticity, though no severe systematic change in variance is

observed. Finally, the residuals-versus-leverage plot confirms the absence of highly influential outliers, as all points lie well below the Cook's Distance reference curves.

We also addressed structural multicollinearity caused by the interaction terms. In the initial model, the interaction produced extremely large Variance Inflation Factors (VIFs) because the product term was inherently correlated with its uncentered components. To resolve this, we mean-centered the predictors involved in the interaction before refitting the model. This adjustment removed the artificial linear dependence, reducing the VIFs for the main predictors to approximately 2.68 and 2.67, and the interaction term to 1.15. These values fall well below the conservative threshold of 5, confirming that multicollinearity has been effectively mitigated and the regression coefficients are stable for inference.

Research Question 2

Is a country's generosity and freedom associated with the likelihood of the country belonging to a region with a higher happiness group, and how do these two factors interact after controlling for GDP, life expectancy, family support, and trust in government?

Model Description

We fit an ordinal logistic regression model with the regional happiness group as the ordered outcome. To construct this variable, we created a regional happiness grouping based on the average happiness score of each geographic region classified by the dataset (e.g., Western Europe, Latin America and the Caribbean, Eastern Asia). For each region, we computed the mean happiness score across its region and classified these regional averages into Low, Medium, and High categories based on tertile cutoffs.

An ordinal model is appropriate because the outcome categories follow a natural ranking, and the proportional-odds framework allows us to estimate shifts toward higher or lower happiness levels. The model relies on three assumptions: (1) a linear relationship between the predictors and the log-odds of being in a higher category; (2) independence of residuals, addressed here with region-clustered robust standard errors to account for shared regional characteristics and repeated measurement; and (3) the proportional-odds assumption, meaning predictor effects are consistent across the category thresholds.

The model incorporated generosity, freedom, and their interaction to assess whether the association between generosity and regional happiness varies across levels of freedom. This interaction is conceptually motivated: generosity may operate differently in contexts with varying degrees of autonomy and social openness, such that its effects may strengthen or weaken depending on institutional conditions. Freedom is also included as a key covariate because it

is plausibly related to both generosity and happiness, and omitting it would risk confounding generosity's independent effect. To account for broader structural determinants of well-being, we additionally controlled for life expectancy, GDP per capita, family support, trust in government, and year.

We also explored including interactions with year (e.g., Freedom × Year, Generosity × Year) to parallel the linear specification in Research Question 1, but these models generated extreme coefficients, near-zero standard errors, and implausibly large test statistics. These issues made the ordinal model unstable and the resulting parameters difficult to interpret. For this reason, we retained year as a main-effect control and excluded its interaction terms, while preserving the theoretically grounded Generosity × Freedom interaction that remained stable and meaningful.

Model Results