A Hierarchical Bayesian Analysis of Global Happiness

Introduction

We analyze global happiness using data from the Cantril Ladder survey, which records self-reported happiness on a scale from 0 (maximum sorrow) to 10 (maximum bliss). These data are combined with country-level indicators of economic wealth (log GDP) and health (life expectancy) to study global trends, country rankings, and continental differences.

To account for the nested structure of the data—yearly observations within countries, and countries within continents—we fit a single hierarchical Bayesian generalized linear model (GLM), enabling inference at global, continental, and country levels.

Data

Two datasets were used: (i) **Cantril Ladder**, containing yearly happiness scores for countries (2011–2024) with continent labels, and (ii) **Countries**, providing country-level covariates (log GDP and life expectancy). The datasets were merged by country; rows with missing covariates were removed. Year, log GDP, and life expectancy were standardized. Countries and continents were encoded as integer indices for hierarchical modeling.

Model

We model happiness scores using a hierarchical Bayesian linear model with three levels: global, continent, and country.

Observation level

For observation i (country–year),

$$y_i \sim \mathcal{N}(\mu_i, \sigma), \tag{1}$$

$$\mu_i = \alpha_{c[i]} + \beta^{(t)}_{k[i]}\, t_i + \beta^{(g)}_{k[i]}\, g_i + \beta^{(l)}_{k[i]}\, l_i, \tag{2}$$

where t_i, g_i, and l_i denote standardized year, log GDP, and life expectancy.

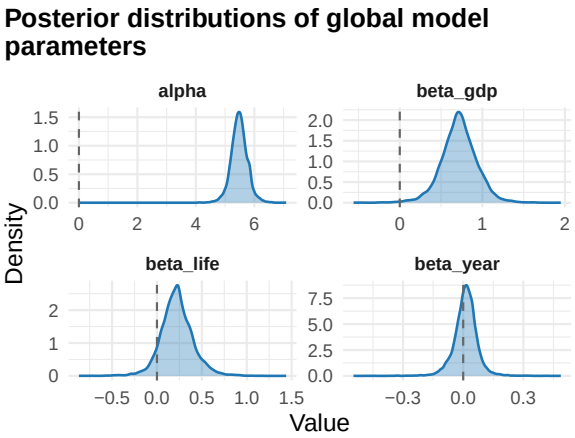

Figure 1: Posterior distributions of global parameters.

Country and continent levels

Country-specific intercepts capture average happiness within each country and follow

$$\alpha_c \sim \mathcal{N}(\alpha_k, \sigma^{(c)}_\alpha),$$

allowing partial pooling toward continent-level means. Continent-level parameters follow

$$\alpha_k \sim \mathcal{N}(\alpha_0, \sigma^{(k)}_\alpha), \tag{3}$$

$$\beta^{(t)}_k \sim \mathcal{N}(\beta^{(t)}_0, \sigma_t), \tag{4}$$

$$\beta^{(g)}_k \sim \mathcal{N}(\beta^{(g)}_0, \sigma_g), \tag{5}$$

$$\beta^{(l)}_k \sim \mathcal{N}(\beta^{(l)}_0, \sigma_l). \tag{6}$$

Global priors

Global parameters were assigned weakly informative priors:

$$\alpha_0 \sim \mathcal{N}(5, 3), \tag{7}$$

$$\beta^{(t)}_0,\ \beta^{(g)}_0,\ \beta^{(l)}_0 \sim \mathcal{N}(0, 1), \tag{8}$$

$$\sigma,\ \sigma^{(c)}_\alpha,\ \sigma^{(k)}_\alpha,\ \sigma_t,\ \sigma_g,\ \sigma_l \sim \text{Exponential}(1). \tag{9}$$

The model was fitted in Stan using 4 chains and 5000 iterations per chain (2500 warmup).

Results

All key parameters show excellent convergence, with $\hat{R}$ values below 1.01 and large effective sample sizes (about 1,800 to 7,000 for global parameters). Some divergent transitions occurred after warmup, reflecting the complexity of the hierarchical variance structure. Nevertheless, posterior summaries of global and continental effects were stable across chains and across repeated runs.

Q1: Global trend over time

The global year slope has posterior mean 0.01 with 95% credible interval $[-0.11, 0.12]$, and $P(\beta_{\text{year,global}} > 0) = 0.62$. Model-implied changes in expected happiness over time are small relative to residual variability. **Conclusion:** There is no strong evidence for a systematic increase in global happiness over time.

Q2: Are humans globally happy?

The global intercept has posterior mean 5.48 with 95% credible interval $[4.91, 6.06]$, and $P(\alpha_{\text{global}} > 5) = 0.96$. **Conclusion:** On average, global happiness is slightly above the neutral midpoint.

Q3: Slovenia's happiness ranking

Posterior country-level intercepts were used to compute ranks. Slovenia's posterior rank has median ≈ 121 with 95% interval $[82, 148]$, and zero posterior probability of being among the top 20 happiest countries. **Conclusion:** Slovenia is very unlikely to rank among the world's happiest countries.

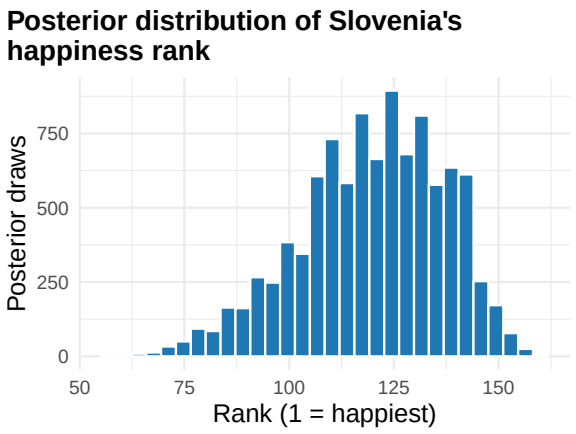

Figure 2: Posterior distribution of Slovenia's happiness rank.

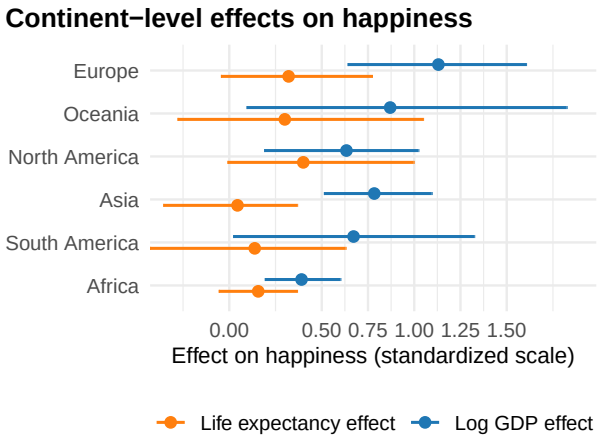

Figure 3: Continent-level effects of GDP and life expectancy, with posterior means and 95% credible intervals.

Q4: Continental differences in health and wealth effects

Substantial heterogeneity is observed across continents. The GDP–happiness relationship is strongest in Europe and weakest in Africa, with $P(\beta_{\text{gdp,Europe}} > \beta_{\text{gdp,Africa}}) = 0.996$. Life expectancy effects are strongest in Europe and North America and weakest in Asia. **Conclusion:** The relationship between happiness, wealth, and health varies markedly by continent.

Discussion

This hierarchical Bayesian analysis enables coherent inference at global, continental, and country levels within a single model. Partial pooling stabilizes estimates, allows probabilistic country rankings, and reveals meaningful regional heterogeneity. Although some divergent transitions indicate challenging posterior geometry—primarily for variance components—global and continental conclusions were stable across runs.

Limitations. Happiness scores are self-reported and may reflect cultural or contextual biases. Country-level covariates are treated as time-invariant, ignoring within-country temporal variation. The linear specification does not capture nonlinearities or interactions. Some variance components exhibit challenging posterior geometry, so uncertainty estimates for those parameters should be interpreted cautiously. Finally, the analysis is observational and does not support causal interpretations.