

The World Happiness Report – 2024

Helliwell, J. F., Layard, R., Sachs, J. D., De Neve, J.-E., Aknin, L. B., & Wang, S. (Eds.). (2024). World Happiness Report 2024. University of Oxford: Wellbeing Research Centre.

<https://happiness-report.s3.amazonaws.com/2024/Ch2+Appendix.pdf>

Summary of Variable Meanings:

1. Happiness Score (Ladder) – Variable ("Ladder")

- How was it collected? From the Gallup World Poll between 2005-2023.
- How is it measured? People were asked about their life ladder (Cantril ladder), where:
 - 0 = the worst life
 - 10 = the best life
- What does it mean? A higher score = greater life satisfaction.
- How to use it? A continuous variable can be analyzed for its relationship with other factors using linear regression.

2. GDP per capita – Variable ("GDP per capita")

- How was it collected? Based on World Bank (WDI) and Penn World Table data.
- How is it calculated? In PPP (Purchasing Power Parity), which adjusts for cost of living differences between countries.
- What does it mean? The higher the value, the wealthier the country.
- Important note: Data for 2023 is forecasted, not direct data, so this should be considered when analyzing.

3. Healthy Life Expectancy (HLE) – Variable ("HLE")

- How was it collected? From the World Health Organization (WHO), with extrapolation for missing years.
- What does it mean? A higher score = healthier life expectancy in the country.
- Are there limitations? The data is not directly updated for 2024 but is based on projections.

4. Social Support – Variable ("social_support")

- How was it collected? Through the question:
 - "If you were in trouble, do you have friends or family members you could rely on?"
- How is it measured? The average of binary answers: 0 = no support, 1 = support available.
- How to interpret? A higher value = more social support.

5. Freedom to Make Life Choices – Variable ("freedom")

- How was it collected? Through the question:
 - "Are you satisfied with the freedom to choose what to do with your life?" (Satisfied = 1, Not satisfied = 0)
 - The average was taken, with values closer to 1 indicating more freedom.

- How to interpret? A higher score = greater personal freedom in the country.

6. Generosity – Variable ("generosity")

- How was it collected? People were asked:
 - "Did you donate money to charity in the past month?"
- How is it calculated? The residual of a regression between generosity and GDP, meaning generosity is measured relative to the country's wealth.

7. Corruption Perception – Variable ("corruption")

- How was it collected? The average of answers to the questions:
 - "Is there widespread corruption in the government?"
 - "Is there corruption in businesses?"
- How to interpret? A higher score = higher perceived corruption, which is negative.

Part A

Question 1:

I chose the dataset from the World Happiness Report 2024 (the most up-to-date) for several reasons.

First, we all want to be happy. We try to navigate our way through life and make decisions that will lead to happiness (or to avoid pain and seek pleasure), but generally, happiness is something that matters to us and is influenced by many factors, which is why it interests and engages me.

Additionally, I have come across numerous articles that highlight Israel among the top-ranking countries. I was curious to understand what could place Israel so high in 2024, especially after experiencing such a difficult war.

Israel ranks fifth in the World Happiness Index, which places it alongside countries that have not experienced war for thousands of years.

Furthermore, I have been exposed multiple times to the narrative that happiness levels tend to be higher in less developed countries. This led me to wonder what could be the factors behind this, whether it is even true, and what generally affects our happiness.

Question 2:

Distribution of All Numeric Variables

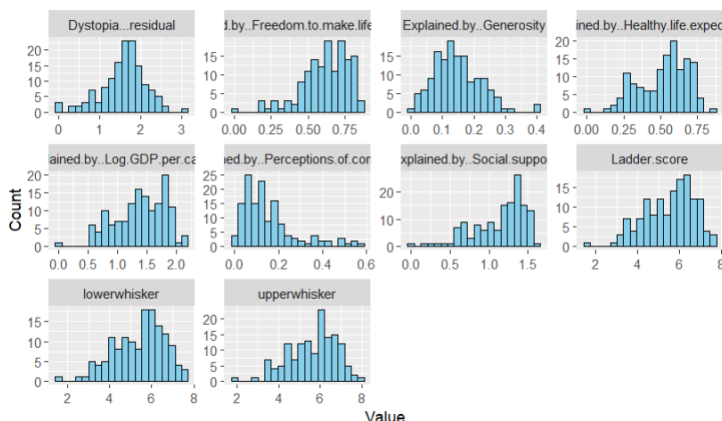


Fig 1. This figure presents the distribution of all numeric variables in the dataset. Each histogram visualizes the frequency of different values for a given variable, providing insights into their spread and central tendency.



Fig 2. This figure presents the distribution of happiness scores across countries. The x-axis represents the "Ladder Score" as a happiness measure, while the y-axis shows the number of countries that selected each score as their national happiness indicator. The histogram, with an overlaid density curve, suggests that most countries report moderate to high happiness levels, peaking around a score of 6.

Question 3:

Defining the research question

Upon reviewing the data, I noticed an interesting point.

Although corruption perception was measured using the average of two questions sent to citizens:

1. Do you think there is corruption in the country?
2. Do you think there is corruption in your workplace?

Where a "Yes" (corruption) received a score of 1, and a "No" (no corruption) received a score of 0, with the average being calculated for all citizens in a given country to yield a number. The result shows that the higher the number, the higher the perception of corruption (the closer the value is to 1, the more corruption is perceived).

However, it appears that in highly developed countries, many of which have democratic regimes and freedom of expression, the corruption perception score tends to be higher, in contrast to less developed countries, which have lower scores on the corruption index.

The dataset does not include variables for freedom of expression or the type of regime. It would be interesting to understand what factors could influence a high corruption perception score and how it might impact the World Happiness Index.

Two Research Questions:

Does the relationship between freedom and perceived corruption differ between developed and developing countries?

How does perceived corruption affect national happiness, and does this effect vary depending on a country's freedom?

Part B+C:

First Research Question:

Does the relationship between freedom and perceived corruption differ between developed and developing countries?

Predictor Variable X: Corruption

Predictor Variables Y: Freedom, Country Development Level (Developed or Not)

Interaction Variable: Freedom * Country Development Level

A multiple regression analysis will be performed.

Results:

A linear regression was conducted to examine the effect of freedom level (freedom) and country development level (developed) on corruption level (corruption), while also considering their interaction.

Main Effect of Freedom

The main effect of freedom was found to be non-significant ($b = 0.030$, $SE = 0.051$, $t(136) = 0.59$, $p = 0.558$, 95% CI $[-0.071, 0.131]$), indicating that when the development level is held constant, there is no evidence that freedom level directly impacts corruption.

Main Effect of Development

The main effect of development was significant ($b = -0.305$, $SE = 0.108$, $t(136) = -2.82$, $p = 0.006$, 95% CI $[-0.518, -0.093]$), indicating that when the freedom level is 0, developed countries exhibit significantly lower levels of corruption compared to developing countries.

Interaction Effect

A significant interaction effect was found between freedom × development ($b = 0.694$, $SE = 0.153$, $t(136) = 4.54$, $p < 0.001$, 95% CI $[0.391, 0.997]$), suggesting that the effect of freedom level on corruption level depends on the development level of the country. Specifically, in developed countries, an increase in freedom level is associated with a sharper rise in corruption compared to developing countries.

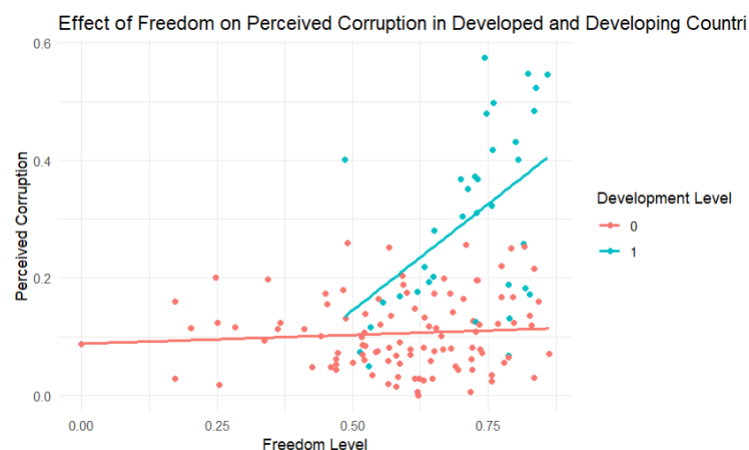


Fig 3. This scatter plot illustrates the effect of freedom level on perceived corruption across developed and developing countries. The interaction between freedom and development level is evident: in developing countries (red), freedom has little impact on perceived corruption,

while in developed countries (blue), higher freedom is associated with a significant increase in perceived corruption.

Second research question:

How does perceived corruption affect national happiness, and does this effect vary depending on a country's freedom?

Predictor Variable X: Happiness (above 5.5 - Happy, below 5.5 - Not Happy)

Predictor Variables Y: Corruption Perception, Freedom

Interaction Variable: Corruption Perception * Freedom

A logistic regression analysis will be performed.

Results:

A logistic regression analysis was conducted to examine how corruption perception (corruption) influences the likelihood of a country being happy (happy_country), and whether this effect varies depending on the level of freedom in the country (freedom).

Main Effects

Corruption Perception: No significant relationship was found between corruption perception and the likelihood of a country being happy (OR = 1.43e-09, 95% CI [3.26e-25, 9137.22], $p = .224$). This suggests that, when the level of freedom in the country is 0, there is no evidence that corruption alone affects the chances of a country being happy.

Freedom: It was found that every one-unit increase in the level of freedom in a country is associated with a 415.18-fold increase in the odds of the country being happy (OR = 415.18, 95% CI [0.82, 403,047.78]), although the effect is not statistically significant ($p = .073$).

Interaction Effect

A positive interaction between corruption perception and the level of freedom in the country was found, but it is not statistically significant (OR = 4.88e+14, 95% CI [6.67e-05, 1.39e+38], $p = .175$). This suggests that there is no significant evidence that the relationship between corruption and happiness changes substantially depending on the level of freedom in the country.

Odds of National Happiness as a Function of Perceived Corruption, with Freedom Level Indicator

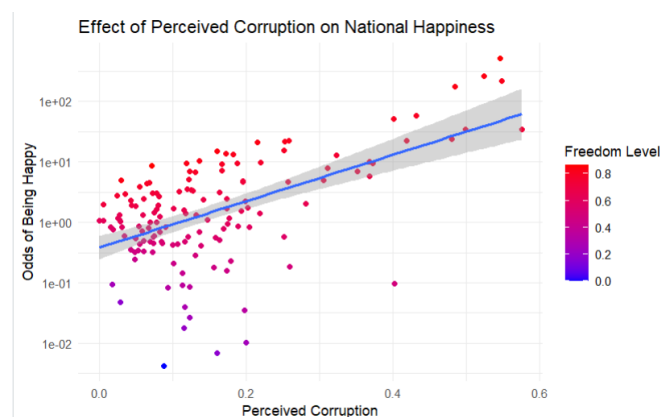


Fig 4. shows the relationship between perceived corruption and happiness odds. The regression line indicates a positive correlation, with data points color-coded by freedom level (red = high, blue = low).

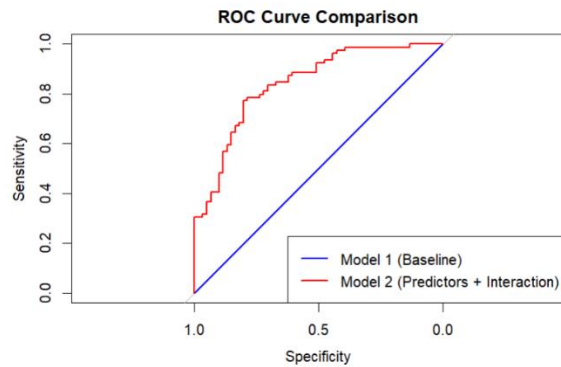


Fig 5. presents an ROC curve comparing two models: a baseline model (blue) and an improved model incorporating predictors and interactions (red). The dependent variable is the classification outcome, while the independent variable represents the model's predicted probabilities. The red curve demonstrates better performance, with an AUC value of **0.84**, indicating improved discrimination ability compared to the baseline model.