# GLOBAL HAPPINESS ANALYSIS: EXPLORING THE ROLE OF SOCIAL SUPPORT AND TRUST (2015–2019)

Nico Bui

17.06.2025

# Contents

# Data Source

**Dataset:** World Happiness Report (2015–2017)
**Source:** Gallup World Poll, hosted publicly on Kaggle
https://www.kaggle.com/datasets/unsdsn/world-happiness
**License:** Free for educational and non-commercial use

## What this data is

This dataset comes from Gallup's global survey, where people in over 150 countries rate their current life satisfaction on a scale from 0 (worst possible life) to 10 (best possible life). The dataset includes both overall happiness scores and a breakdown of the main factors that contribute to those scores.

The factors include:

- Economic production (GDP per capita)

- Social support (listed as "Family")

- Life expectancy (Health)

- Freedom to make life choices

- Generosity

- Trust in government and institutions

- A residual component showing unexplained variance

All of this is compiled into a single dataset with one row per country, along with that country's rank and region.

**Why I chose it**

I wanted a dataset that reflects real-world conditions and emotional well-being in a way that is still structured, measurable, and open. This dataset is:

- Clean and already aggregated

- Easy to understand

- Widely used and published by trusted sources (Gallup, UN)

- Relevant to social and psychological dimensions of quality of life

Even though it doesn't include individual-level behavior or local event data, it still provides a strong foundation to explore questions about what drives happiness and how countries differ in that regard.

## How the data was collected

The happiness scores are based on a survey question where respondents rate their lives on a scale from 0 to 10 (known as the Cantril ladder).
The contributing factors are based on standardized indicators collected from various international data sources (World Bank, WHO, etc.).
Gallup applies weights to ensure national representation, and all results are averaged across multiple years for consistency.

# What's in the dataset

- Country and region

- Happiness score and rank

- Component scores for each happiness factor (GDP, Family, Health, Freedom, etc.)

- Dystopia residual (unexplained happiness)

Each row represents one country. All key variables are numeric, making the dataset suitable for comparison, correlation analysis, and clustering.

# Data Profile

The dataset contains separate files for each year from 2015 to 2019. While the overall structure is consistent, country-level happiness scores plus contributing factors, there are differences in both column names and included variables across years.

- 2015–2017 contain more granular columns, such as confidence intervals and a Dystopia Residual.

- 2018–2019 simplify the dataset to fewer variables and rename several columns (Family becomes Social support).

- Columns like Standard Error and Confidence Interval are dropped in later years.

- Some column names also use different punctuation or casing (Happiness Score vs. Score, GDP per capita vs. Economy, GDP per Capita).

## Variable Overview

| Variable | Description |
| --- | --- |
| Country | Name of the country |
| Year | Year of the report (2015 to 2019) |
| Happiness Score | Overall happiness rating (0–10 scale) based on Gallup World Poll |
| GDP | GDP per capita |
| Social Support | Availability of social support from friends/family |
| Life Expectancy | Healthy life expectancy |
| Freedom | Freedom to make life choices |
| Generosity | Willingness to donate and help others |
| Trust | Perception of corruption in government/business |

## Descriptive Stats Analysis

- The mean happiness score is ~5.38, ranging from 2.69 (lowest) to 7.77 (highest).

- GDP per capita ranges from 0 to 2.096, indicating large income differences.

- Social support and life expectancy are positively correlated with happiness.

- Trust in government is generally low (mean = 0.125).

|  | Year | Happiness Score | GDP | Social Support | Life Expectancy | Freedom | Generosity | Trust |
|---|---|---|---|---|---|---|---|---|
| count | 782.000000 | 782.000000 | 782.000000 | 782.000000 | 782.000000 | 782.000000 | 782.000000 | 782.000000 |
| mean | 2016.993606 | 5.379018 | 0.916047 | 1.078392 | 0.612416 | 0.411091 | 0.218576 | 0.125418 |
| std | 1.417364 | 1.127456 | 0.407340 | 0.329548 | 0.248309 | 0.152880 | 0.122321 | 0.105750 |
| min | 2015.000000 | 2.693000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 25% | 2016.000000 | 4.509750 | 0.606500 | 0.869363 | 0.440183 | 0.309768 | 0.130000 | 0.054250 |
| 50% | 2017.000000 | 5.322000 | 0.982205 | 1.124735 | 0.647310 | 0.431000 | 0.201982 | 0.091033 |
| 75% | 2018.000000 | 6.189500 | 1.236187 | 1.327250 | 0.808000 | 0.531000 | 0.278832 | 0.155861 |
| max | 2019.000000 | 7.769000 | 2.096000 | 1.644000 | 1.141000 | 0.724000 | 0.838075 | 0.551910 |

## Sample size

There are 782 rows across 5 years, covering approximately 155 countries per year.

| Year | # Columns | Key Differences |
|---|---|---|
| 2015 | 12 | Includes Standard Error and Dystopia Residual |
| 2016 | 13 | Adds Confidence Intervals (Lower + Upper) |
| 2017 | 12 | Renamed columns with dots (Economy..GDP.per.Capita.), includes Whisker.low/high |
| 2018 | 9 | Drops confidence intervals and residuals, renames everything |
| 2019 | 9 | Same as 2018 |

I will keep only the common variables across all years: Country, Year, Happiness Score, GDP, Social Support, Life Expectancy, Freedom, Generosity, Trust.

Only one missing value was found (Trust, UAE 2018), which was filled using the average trust score of that year to preserve completeness. Country names were also standardized to ensure consistent grouping across years.

## Hypothesis Test

Impact of Social Support on Happiness
Null Hypothesis ($H_0$): There is no significant difference in happiness scores between countries with high and low social support.
Alternative Hypothesis ($H_1$): Countries with higher social support report significantly different higher happiness scores.

- **Test**: Two-sample t-test assuming unequal variances
  Significance level (α): 0.05
- Results:
  - t-Statistic how big the difference is between groups: 16.53
  - p-Value (Two-Tailed) how likely it is this difference happened by chance: $8.04 \times 10^{-53}$
  - t-Critical (Two-Tailed): 1.96
- Interpretation:
  The p-value is much smaller than 0.05.
  → Reject the null hypothesis.

- Conclusion:
  There is a statistically significant difference in happiness scores between countries with high and low levels of social support.

## Limitations

- The data shows only country averages, not differences between people within each country.

- Each year used different column names, so I had to clean and standardize them manually.

- Some values were missing (e.g. "Trust" for one country) — I filled them using the yearly average.

- The dataset only covers 2015 to 2019, so it doesn't reflect recent events like COVID-19 or wars.

## Ethical Considerations

- The results show correlation, not causation which means we can't say for sure that one thing causes another.

- Survey answers can be influenced by culture so some people may rate their happiness higher or lower based on how they were raised.

- The data is public and anonymous, so there are no privacy issues.

- This data could be used in policymaking so it's important to be careful with interpretation to avoid unfair comparisons.

# Questions to explore

- Is there a positive correlation between social support and happiness score?

- Which countries have high social support but low happiness and vice versa?

- Among GDP, social support, life expectancy, freedom, generosity, and trust, which one has the strongest correlation with happiness?

- Are there countries with low GDP but high happiness?

- Did the importance of social support or trust change from 2015 to 2019?

- Which countries had the lowest scores in social support or trust?

- Which countries were the top 10 happiest each year (2015–2019)?

- How does happiness relate to social support across countries?

- Which countries gained or lost the most happiness between 2015 and 2019?

- How does average happiness differ between countries with high vs. low trust in government?

- Which factor (GDP, trust, generosity, etc.) is most strongly related to happiness?