

World Happiness Report 2015-2019

Catherine Walerko

12/21/22



Data: [World Happiness 2015-2019](#)

Data Source

The World Happiness Report surveys the state of global happiness by ranking 155 countries by their happiness levels. Respondents to the World Gallup Poll were asked to rank their current lives on a scale of zero to ten, with ten representing the best possible life. Six factors – economic production, social support, life expectancy, freedom, absence of corruption and generosity – are also estimated as to the impact they may have on the life evaluations in each country when compared to the fictional country of Dystopia. Dystopia serves as a benchmark with values equal to the lowest national averages for all six factors.

This is an external data source with data being manually collected annually via survey for Gallup World Poll. As a well-respected and global source, this data can be considered trustworthy.

As someone who loves to travel and would have dreamed of working abroad, I chose this data set to learn where and why people are – or aren't – happy with where they live.

Data Profile

Data cleaned in Excel prior to merging in Python.

Please see Jupyter Notebook 6.1 for further Data Profiling.

2015 Data Set:

- Removed "Region" and "Standard Error" columns to keep consistent with following years.
- Columns renamed for clarity
- No blanks or duplicates found
- Added "Year" column

2016 Data Set:

- Removed "Region", "Lower Confidence Interval," and "Upper Confidence Interval" columns to keep consistent with following years.
- Columns renamed for clarity
- No blanks or duplicates found

- Added “Year” column

2017 Data Set:

- Removed “whisker.high” and “whisker.low” columns to keep consistent with following years.
- Columns renamed for clarity
- No blanks or duplicates found
- Added “Year” column

2018 Data Set:

- Columns renamed for clarity
- Columns reordered for consistency
- “Dystopia_Residuals” column calculated and added for consistency
- 1 N/A found (Govt_Corruption for United Arab Emirates) and changed to 0
- No duplicates found
- Added “Year” column

2019 Data Set:

- Columns renamed for clarity
- Columns reordered for consistency
- “Dystopia_Residuals” column calculated and added for consistency
- No blanks or duplicates found
- Added “Year” column

Column	Column Description	Data Type	Time Variant
Year	Report Year	Quantitative, Discrete	Yes
Country	Country	Qualitative, Nominal	No
Happiness_Rank	Country’s Happiness Rank Based on Happiness Score	Quantitative, Discrete	No
Happiness_Score	Country’s Happiness Score Averaged from Survey Responses	Quantitative, Continuous	No
GDP_Capita	Estimated Contribution of Country’s GDP per	Quantitative, Continuous	No

	Capita to Happiness Score		
Family	Estimated Contribution of Country's Social Support to Happiness Score	Quantitative, Continuous	No
Life_Expectancy	Estimated Contribution of Country's Life Expectancy to Happiness Score	Quantitative, Continuous	NO
Freedom	Estimated Contribution of Country's Amount of Freedom to Happiness Score	Quantitative, Continuous	No
Generosity	Estimated Contribution of Country's Level of Generosity to Happiness Score	Quantitative, Continuous	No
Govt_Corruption	Estimated Contribution of Country's Perception of Government Corruption to Happiness Score	Quantitative, Continuous	No
Dystopia_Residual	Dystopia Happiness Score (1.85) + the Residual or Unexplained Value for Each Country	Quantitative, Continuous	No

Limitations, Ethics, and Methodology

- Merging the data sets only resulted in 782 rows of data – short of the 1500 requested in the project brief.
- The estimated six variables are added up to get the Happiness Score, so the model may be unreliable to predict Happiness Scores.

- Per Gallup's website:
"...Gallup weights World Poll samples to correct for unequal selection probability, nonresponse, and double coverage of landline and cellphone users when using both cellphone and landline frames. Gallup also weights its final samples to match the national demographics of each selected country. The margin of error for each sample reflects the influence of data weighting. In addition to sampling error, question wording and practical difficulties in conducting surveys can introduce error or bias into the findings of public opinion polls."

Source: [Gallup Methodology](#)

- Per Gallup's website:
"With some exceptions, all samples are probability based and nationally representative of the resident population aged 15 and older. The coverage area is the entire country including rural areas, and the sampling frame represents the entire civilian, non-institutionalized adult population of the country. Exceptions include areas where the safety of the interviewing staff is threatened and scarcely populated islands in some countries."

Source: [Gallup Methodology](#)

- Per Gallup's website:
"The typical survey includes at least 1,000 individuals. In some countries, Gallup collects oversamples in major cities or areas of special interest. Additionally, in some large countries, such as China and Russia, sample sizes of at least 2,000 are collected. Although rare, in some instances, the sample size is between 500 and 1,000."

Source: [Gallup Methodology](#)

Questions to Explore

Which of the six variables contribute most to the Happiness Score of the top 10 countries?

Which of the six variables contribute the least to the Happiness Score of the bottom 10 countries?

Which country or countries have gained or lost the most in rankings and what variable(s) is/are estimated to be the cause of the shift?

What is the relationship between GDP per capita and the perception of government corruption?