

# World Happiness Index Analysis 2023 Data Information

## Data Source

### Data Sourcing

This is an external data set owned by the Sustainable Development Network. The data comes from the World Happiness Report, a yearly survey on global happiness that influences policy-making decisions, and thus is susceptible to some bias. Political influences may sway respondents to lie on survey. To mitigate some of this bias, the sample is “national representative” and actual score is analyzed other parties. Thus, the source is relatively trustworthy. The data can be accessed [here](#).

### Data Collection

This data is externally collected via voluntary surveys once a year from 2012 to the present day. The data used for this analysis is from 2015 to 2019.

### Data Contents

The World Happiness Report (WHR) uses survey data from the Gallup World Poll. The scores in the WHR are based on answers to the Cantril ladder, which asks respondents to rate aspects of their lives on a scale of one to ten (ten being the best). Then, Gallup weights are applied to make the estimates for the happiness score. Further estimates are used to see how much economic production, social support, life expectancy, freedom, absence of corruption, and generosity contribute to the happiness score.

### Limitations and Ethics

Due to the nature of the “national representative” sample, the data may have sample bias. The sample may reflect the population in a way that does not represent marginalized groups. It may also have some exclusion bias depending on which countries are included in the survey. For example, if many western European countries abstained from participating, an analysis over the whole of Europe may have inaccurate results.

The data is also limited by the grain of results. Happiness may vary from city to city, but the data only looks at the entire country.

### Relevancy

The analysis aims to see which factors influence human happiness. Since the data cover the World Happiness Report, it is extremely relevant.

## Data Profile

### Data Cleaning

The following table details the list of wrangling and cleaning procedures done.

*Table 1. Summary of wrangling, cleaning, and merging procedures for the World Happiness Index data sets*

Action	Data Set	Reason	Data Set Size
Dropped ‘Standard Error’, ‘Dystopia Residual,’ and ‘Region’ columns	2015 World Happiness Index	They were not present all data set – not compatible with merge	(158, 9)

Dropped 'Region', 'Lower Confidence Interval', 'Upper Confidence Interval', and 'Dystopia Residual' columns	2016 World Happiness Index	They were not present all data set – not compatible with merge	(157, 9)
Dropped 'Whisker.low', 'Whisker.high', and 'Dystopia.Residual' columns	2017 World Happiness Index	They were not present all data set – not compatible with merge	(155, 9)
Renamed: <ul style="list-style-type: none"> <li>'Country' to 'location',</li> <li>'Happiness Rank' to 'rank',</li> <li>'Happiness Score' to 'happiness_score',</li> <li>'Economy (GDP per Capita)' to 'gdp_per_capita',</li> <li>'Family' to 'social_support',</li> <li>'Health (Life Expectancy)' to 'life_expectancy',</li> <li>'Freedom' to 'freedom',</li> <li>'Trust (Government Corruption)' to 'trust_in_government',</li> <li>'Generosity' to 'generosity'</li> </ul>	2015 World Happiness Index	Column consistency for the merge	Unchanged (158, 9)
Renamed: <ul style="list-style-type: none"> <li>'Country' to 'location',</li> <li>'Happiness Rank' to 'rank',</li> <li>'Happiness Score' to 'happiness_score',</li> <li>'Economy (GDP per Capita)' to 'gdp_per_capita',</li> <li>'Family' to 'social_support',</li> <li>'Health (Life Expectancy)' to 'life_expectancy',</li> <li>'Freedom' to 'freedom',</li> <li>'Trust (Government Corruption)' to 'trust_in_government',</li> <li>'Generosity' to 'generosity'</li> </ul>	2016 World Happiness Index	Column consistency for the merge	Unchanged (157, 9)
Renamed: <ul style="list-style-type: none"> <li>'Country' to 'location',</li> <li>'Happiness.Rank' to 'rank',</li> <li>'Happiness.Score' to 'happiness_score',</li> <li>'Economy..GDP.per.Capita.' to 'gdp_per_capita',</li> <li>'Family' to 'social_support',</li> <li>'Health..Life.Expectancy.' to 'life_expectancy',</li> <li>'Freedom' to 'freedom',</li> <li>'Trust..Government.Corruption.' to 'trust_in_government',</li> <li>'Generosity' to 'generosity'</li> </ul>	2017 World Happiness Index	Column consistency for the merge	Unchanged (155, 9)
Renamed: <ul style="list-style-type: none"> <li>'Country or region' to 'location',</li> <li>'Overall rank' to 'rank',</li> <li>'Score' to 'happiness_score',</li> </ul>	2018 World Happiness Index	Column consistency for the merge	Unchanged (156, 9)

<ul style="list-style-type: none"> <li>• 'GDP per capita' to 'gdp_per_capita',</li> <li>• 'Social support' to 'social_support',</li> <li>• 'Healthy life expectancy' to 'life_expectancy',</li> <li>• 'Freedom to make life choices' to 'freedom',</li> <li>• 'Perceptions of corruption' to 'trust_in_government',</li> <li>• 'Generosity' to 'generosity'</li> </ul>			
Renamed: <ul style="list-style-type: none"> <li>• 'Country or region' to 'location',</li> <li>• 'Overall rank' to 'rank',</li> <li>• 'Score' to 'happiness_score',</li> <li>• 'GDP per capita' to 'gdp_per_capita',</li> <li>• 'Social support' to 'social_support',</li> <li>• 'Healthy life expectancy' to 'life_expectancy',</li> <li>• 'Freedom to make life choices' to 'freedom',</li> <li>• 'Perceptions of corruption' to 'trust_in_government',</li> <li>• 'Generosity' to 'generosity'</li> </ul>	2019 World Happiness Index	Column consistency for the merge	Unchanged (156, 9)
Added year column with corresponding year	2015 World Happiness Index 2016 World Happiness Index 2017 World Happiness Index 2018 World Happiness Index 2019 World Happiness Index	Needed for data to be distinguishable after the merge	(158, 10) (157, 10) (155, 10) (156, 10) (156, 10)
Merged all data sets	2015 World Happiness Index 2016 World Happiness Index 2017 World Happiness Index 2018 World Happiness Index 2019 World Happiness Index	To allow for further analysis	(782, 10)
Dropped one row missing trust_in_government data	Merged 2015 – 2019 World Happiness Index	Was less than 5% of the data set and missing important information	(781, 10)

## Basic Descriptive Statistics

Table 2. Data types in the World Happiness Index data set

Variable	Relationship to time	Structure	Characteristic	Characteristic Type
----------	----------------------	-----------	----------------	---------------------

location	invariant	structured	qualitative	nominal
rank	variant	structured	qualitative	ordinal
happiness_score	variant	structured	quantitative	continuous
gdp_per_capita	variant	structured	quantitative	continuous
social_support	variant	structured	quantitative	continuous
life_expectancy	variant	structured	quantitative	continuous
freedom	variant	structured	quantitative	continuous
trust_in_government	variant	structured	quantitative	continuous
generosity	variant	structured	quantitative	continuous
year	variant	structured	qualitative	ordinal

*Table 3. Data integrity of the World Happiness Index data set*

	Minimum	Maximum	Mean	Thoughts
<b>rank</b>	1	158	79	The values are reasonable. At most, 158 countries are ranked in the World Happiness Index.
<b>happiness_score</b>	2.693	7.769	5.378	The average happiness score is only slightly above neutral, but all the values are reasonable.
<b>gdp_per_capita</b>	0	1.871	0.915	The maximum and mean are reasonable, but it seems unlikely that the variable would play no role in happiness. Overall, economics seems to play an important role in happiness.
<b>social_support</b>	0	1.644	1.079	The maximum and mean are reasonable, but it seems unlikely that the variable would play no role in happiness. Overall, social support seems to play an important role in happiness.
<b>life_expectancy</b>	0	1.141	0.612	The maximum and mean are reasonable, but it seems unlikely that the variable would play no role in happiness. Overall, life expectancy seems to play an important role in happiness.
<b>freedom</b>	0	0.724	0.411	The maximum and mean are reasonable, but it seems unlikely that the variable would play no role in

				happiness. Overall, economics seems to play a moderate role in happiness.
trust_in_government	0	0.552	0.125	The maximum and mean are reasonable, but it seems unlikely that the variable would play no role in happiness. Overall, political corruption does not seem to play a large role in happiness
generosity	0	0.838	0.219	The maximum and mean are reasonable, but it seems unlikely that the variable would play no role in happiness. Overall, generosity does not seem to play a large role in happiness
year	2015	2019	2017	The values are reasonable

## Exploration Questions

The following is a list of questions to explore.

- Who is the happiest?
  - Who is the happiest country on average?
  - Who is the least happy country on average?
  - How has happiness changed over time?
  - Is there a trend in happiness?
- What type of policies should lawmakers focus on to ensure a happy population?
  - Are there regional influences on happiness?
  - Have the top influences changed over time?
  - Is the emphasis on economics unique to capitalistic countries?
  - Is it moral to focus on what makes people the “happiest” at the expense of what makes them the healthiest?
  - Are there political influences that cause government corruption to have little impact on happiness?
- What are happy-trait strengths and weaknesses?
  - Are there any aspects that regions are doing well in?
  - Are there any region-specific areas that need improvement?
  - Can regions be grouped by what makes them happy?
- What does the data mean explicitly?
  - What does it mean to live in a “generous” nation?
  - Do different cultural backgrounds impact how respondents view “freedom”?