

# Regional Patterns in Global Happiness

Student Number: Z22590018

Student Name: Bakhtiyor Sohibnazarov

Data Visualisation (CMP-X302-0) Assessment Report

Word Count: 2600

# Table of Contents

Table of Contents .....	ii
1. Introduction & Storytelling Context .....	1
2. Data Preparation.....	2
2.1 Initial Data Structuring and Inspection .....	2
2.2 Temporal Consistency and Country Filtering.....	3
2.3 Handling Missing Data .....	4
3. Exploratory Data Analysis (EDA) .....	5
3.1 Univariate Exploration: First Impressions.....	5
3.2 Bivariate Exploration: Relationships with Happiness .....	6
3.3 Regional Analysis.....	7
4. Research Questions .....	8
5. Visualisation and Analysis.....	9
6. Discussion and Interpretation .....	14
7. Conclusion.....	16
References.....	17
Appendices.....	I
Appendix A: Data Preparation Ouputs.....	II
Appendix B: Colourblind-Friendly Guide .....	III

# 1. Introduction & Storytelling Context

Happiness is frequently overlooked in assessments of societal progress, where countries are usually judged by strict economic metrics. While economic strength matters, it does not fully capture how people experience their lives. Factors such as social support, health, and freedom also shape wellbeing, and their importance varies across the world. [1], [2]

Using the World Happiness dataset, this report takes a data-driven storytelling approach to examine global happiness patterns through visual analysis. Its main aim is to explore how key indicators, including GDP per capita, social support, health, freedom, generosity, and corruption perceptions relate to happiness in different regions.

The analysis begins with data preparation and exploratory review to identify initial trends. These inform specific research questions, which are then examined through tailored visualisations. The report concludes by drawing together insights into a clear narrative, highlighting regional differences, data limitations, and what this means for understanding happiness worldwide.

## 2. Data Preparation

Data preparation was an essential step in this analysis. To ensure reliable results, the dataset was carefully cleaned and standardized. This involved handling missing values thoughtfully to avoid bias, adjusting the data to be comparable across different years, and making sure countries had consistent data over time.

### 2.1 Initial Data Structuring and Inspection

The dataset provides annual country-level observations, with the primary outcome variable being the Life Ladder score, commonly used as a measure of self-reported happiness. This is accompanied by a set of economic and social indicators, including GDP per capita, social support, healthy life expectancy at birth, freedom to make life choices, generosity, and perceptions of corruption. (Appendix A, File A1)

Initial inspection focused on understanding the structure of the dataset, verifying variable types, and assessing overall completeness. Column names were standardised to improve clarity and consistency across the analytical workflow.

Two affect-based variables [3], positive affect and negative affect were excluded from subsequent analysis. While these measures capture short-term emotional states, they are conceptually distinct from the broader evaluative nature of the Life Ladder score, which reflects long-term life satisfaction. Retaining affect variables alongside structural socio-economic indicators risked conflating transient mood fluctuations with more stable determinants of wellbeing. Their removal therefore ensured greater conceptual coherence in the analysis. (Appendix A, File A2)

## 2.2 Temporal Consistency and Country Filtering

The raw dataset provides yearly data by country, which makes it continuous over time and suitable for temporal analysis. Of course, this introduces a new problem that should be handled before analysis. Temporal coverage may not be consistent across the reported years for different countries, meaning that some countries may have reported more or less data over time. This could introduce fabricated bias if averaged without ensuring consistency in temporal coverage. To inspect this issue, a temporal coverage inspection was performed, which helps determine the next steps in the data pre-processing.

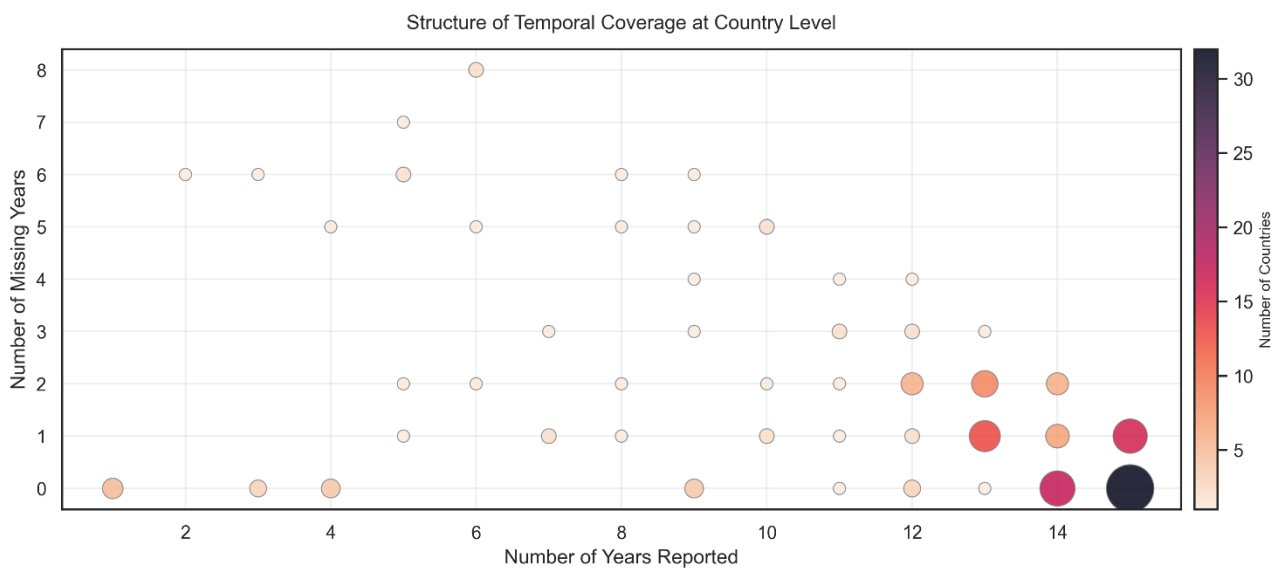


Fig. 1. Bubble Plot of Reported and Missing Years

From the figure (Fig. 1), temporal coverage is inconsistent, while more than 30 countries reported 15 years of consecutive data, some countries reported only one year, and some have very large gaps between reported years, which should not be ignored before analysis. The possible solution is to identify a common time period reported by most countries and lock dataset for that timeframe.

### **What is the minimal temporal coverage that can be both useful in pre-processing and analysis?**

A single year would allow basic comparison but offers limited analytical robustness especially if comparisons are to be made against time, then two years can be enough though it won't provide enough temporal coverage for potential imputations in the future. As a result, 2015–2017-time window has been automatically chosen as this window balances data availability with temporal consistency, retaining 135 out of 166 countries while minimising bias introduced by uneven reporting.

## 2.3 Handling Missing Data

As temporal coverage inconsistency is addressed, missing data still cannot be ignored. Using a three-year window provides sufficient temporal context to impute potential missing values without introducing artificial distortion.

After observation of the missing data, the cleanup process was conducted in three stages.

(Appendix A, File A3)

1. **Column-wise deletion:** The *Corruption* variable was removed entirely due to pervasive and structurally uneven missingness across regions, which precluded reliable imputation or meaningful analysis.
2. **Country-wise exclusion:** South Sudan was identified as an outlier with fundamental data gaps in core economic indicators, particularly GDP, and was removed to maintain column stability.
3. **Controlled interpolation:** For the remaining predictors missing values were interpolated only within countries over time using a linear method. This approach preserves cross-country variation while estimating missing values based on each country's own temporal trajectory. [4], [5], (Appendix A, File A4)

This multi-stage cleaning process resulted in a final analytical dataset of 134 countries, as South Sudan was excluded due to extensive missing data that could not be reliably interpolated. Some missing values remain in the *health* column for three countries; however, these countries were not removed from the dataset because the missingness is limited to a single column and can be addressed during the analysis stage. (Appendix A, File A5, A6)

### 3. Exploratory Data Analysis (EDA)

Before forming specific research questions, an exploratory data analysis [6], [7] was conducted to observe initial trends, distributions, and relationships within the dataset. This process involved examining variables both individually and in combination to understand the data's structure and identify potential patterns worth deeper investigation.

#### 3.1 Univariate Exploration: First Impressions

Univariate analysis focuses on examining one variable at a time to understand its distribution, central tendency, and spread. This step provided foundational insights into how each happiness-related factor behaves independently.

Happiness scores were broadly distributed, with most countries reporting values between 4 and 7 out of 10. Economic and health indicators showed high global inequality, with some nations being far wealthier and healthier than others.

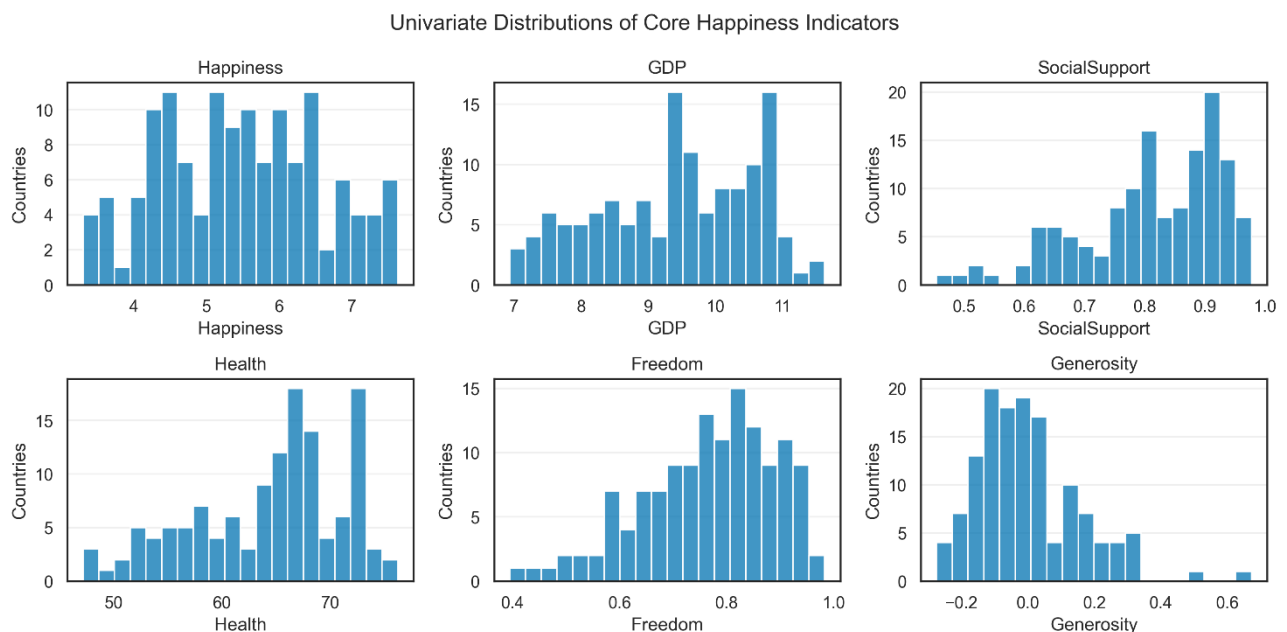


Fig. 2. Univariate Analysis

In contrast, social support and freedom levels remained relatively high even in less affluent regions. Generosity, however, was notably low across most countries, with only a small number of outliers reporting higher values (Fig. 2).

### 3.2 Bivariate Exploration: Relationships with Happiness

As initial observations made, bivariate analysis will be conducted to understand relationships between attributes. Happiness will be chosen as core attribute and relationship of other attributes towards happiness will be examined

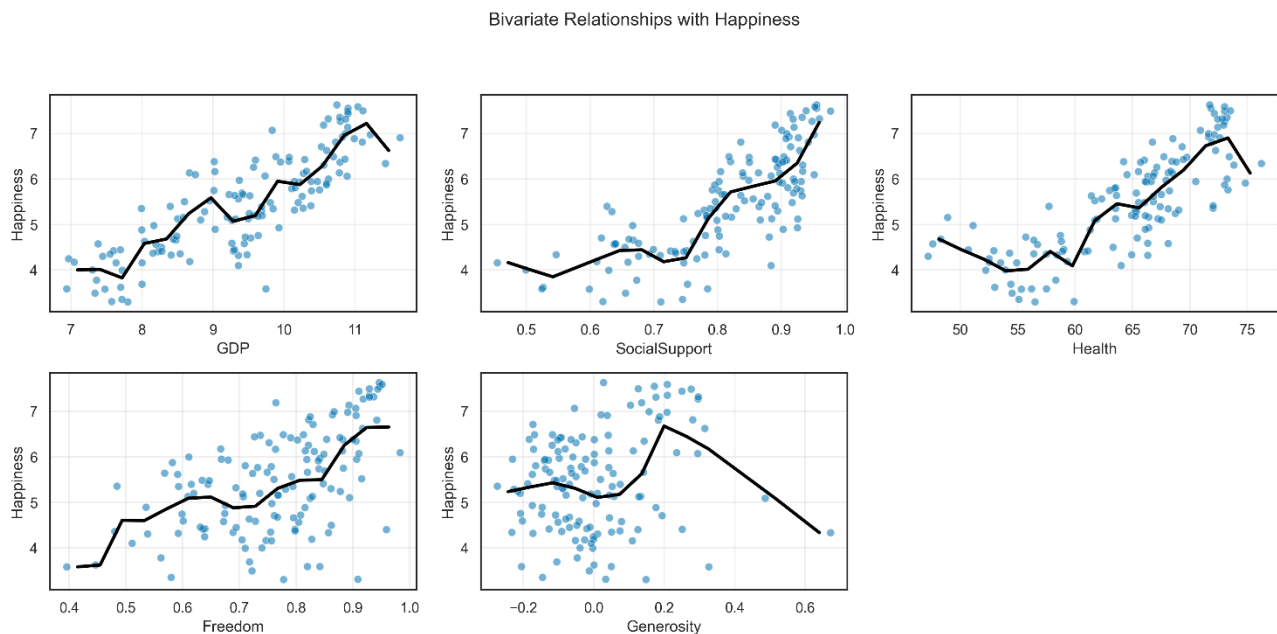


Fig. 3. Bivariate Analysis

Economic indicators such as GDP and Health showed strong positive associations with happiness, although diminishing returns were observed beyond certain levels. Social support exhibited the clearest linear relationship with happiness, while freedom demonstrated a weaker and more dispersed association (Fig. 3). In theory, greater wealth and better health should correspond to higher happiness; however, the observations in this figure show a more dispersed relationship.

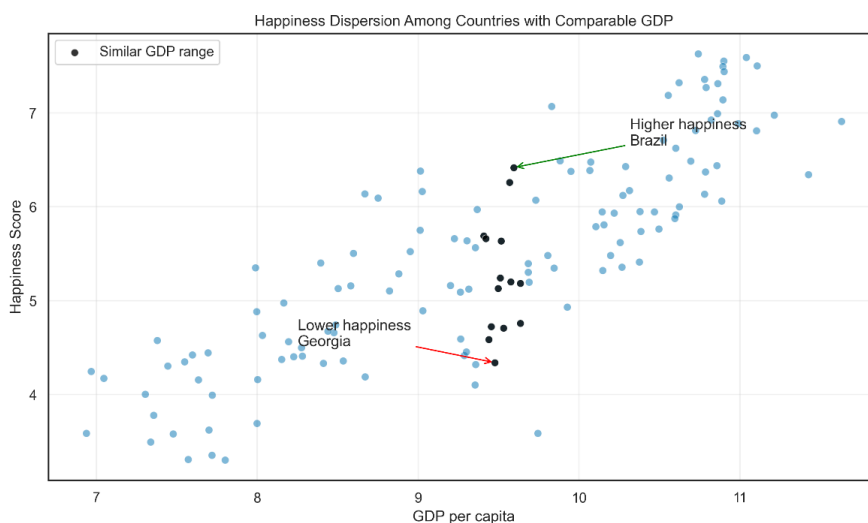


Fig. 4. GDP vs Happiness close inspection

To examine this argument more closely, the relationship between happiness and GDP can be explored in greater detail to better understand the observed dispersion. As shown in this figure, Brazil is considerably happier than Georgia, even though



both countries fall within a similar GDP range. Further research is necessary to understand the influence of non-economic factors on national happiness levels (Fig. 4).

### 3.3 Regional Analysis

To extent small number of countries were split into regions to observe natural clustering by region. European countries consistently reported happiness above what their economic scores predicted, while several Asian and African nations fell below. This early clustering hinted that geography might meaningfully shape how happiness is expressed (Fig. 5).

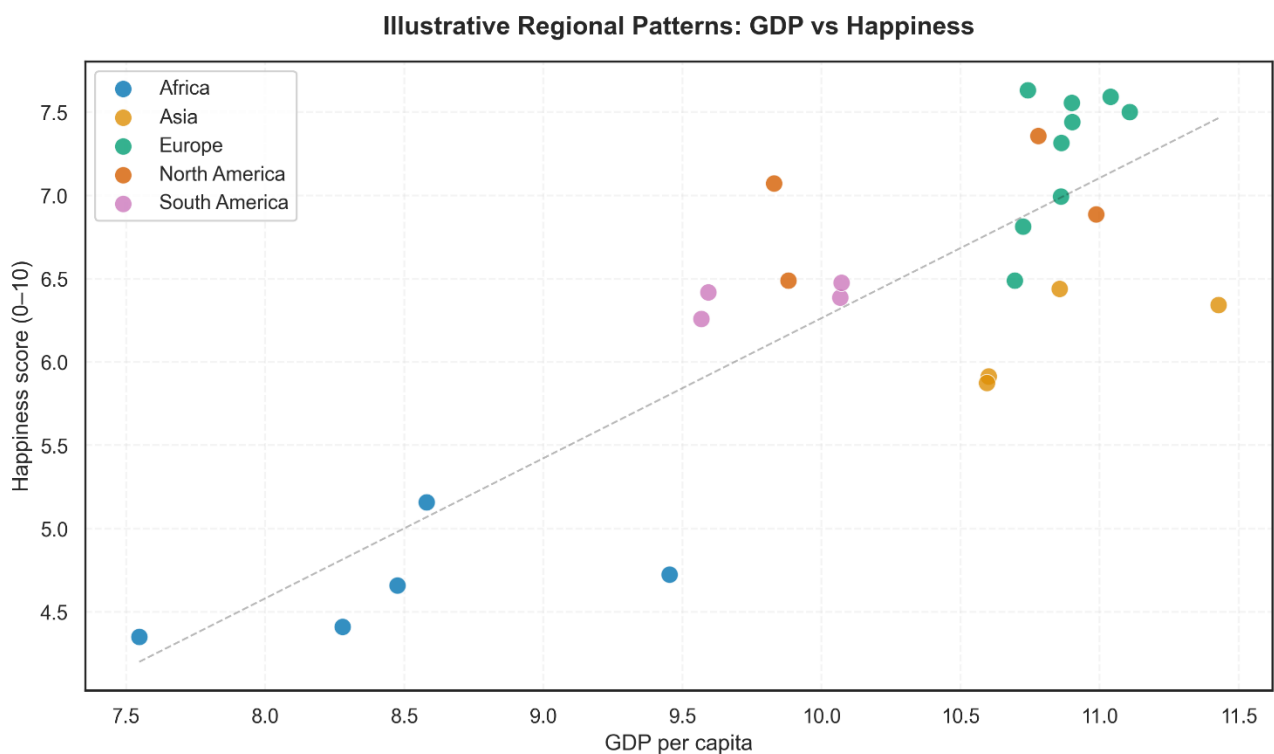


Fig. 5. Regional Natural Clustering

## 4. Research Questions

The exploratory analysis revealed that national happiness cannot be fully explained by economic or social factors alone. Regional clustering and notable happiness differences among economically similar countries suggest that geographic context plays a meaningful role in well-being. These observations motivated the following research questions to examine how regional differences shape global happiness patterns.

### **Research Question 1: Regional Influence Beyond Economics**

**Do countries within the same geographic region exhibit similar happiness levels after controlling for economic and social factors?**

This question builds on the observed regional clustering in exploratory analysis, where European countries tended to exceed the global trend while parts of Asia and Africa fell below it. It tests whether regional context adds explanatory power beyond standard predictors such as GDP and social support.

### **Research Question 2: Regional Variation in Predictor Importance**

**Do key determinants of happiness vary in influence across world regions?**

Although global trends highlight strong predictors like social support, exploratory analysis suggested that the strength of these relationships differs by region. This question examines whether factors such as freedom, health, and generosity contribute unevenly to happiness across regions, adding nuance to global models and supporting region-specific policy insights.

### **Research Question 3: Regional Over- and Under-Performance**

**Do some regions consistently report higher or lower happiness than their economic conditions would predict?**

Observed happiness gaps among economically similar countries motivate an investigation into systematic regional over or under-performance, identifying regions that translate conditions into well-being more or less effectively.

Together, these questions progress from establishing whether regions matter (RQ1), to how their effects differ (RQ2), and finally to identifying regional patterns of over- and under-performance (RQ3), providing a coherent framework for analyzing geographic influences on global happiness.

## 5. Visualisation and Analysis

All visualisations use a colourblind-friendly palette and consistent styling to support comparison across figures. [8], [9], (Appendix B, File B1)

### RQ1: Baseline Model

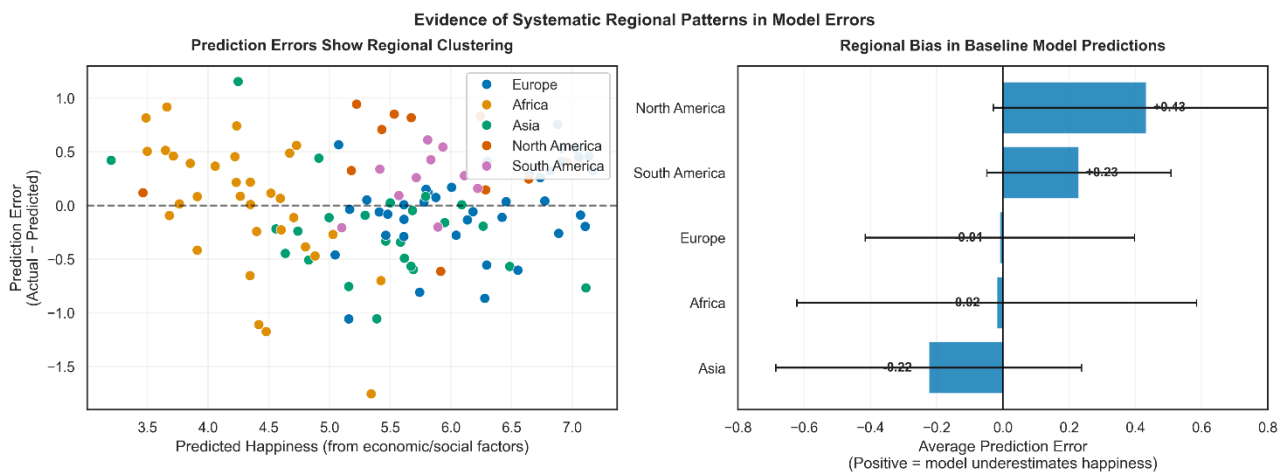


Fig. 6. Baseline Model of Happiness Prediction

This figure displays predicted happiness scores from the baseline linear regression model [10] alongside observed values. The residuals show visible clustering by geographic region, with countries from the same region tending to have similar prediction errors.

### RQ1: Regional Clustering Prediction

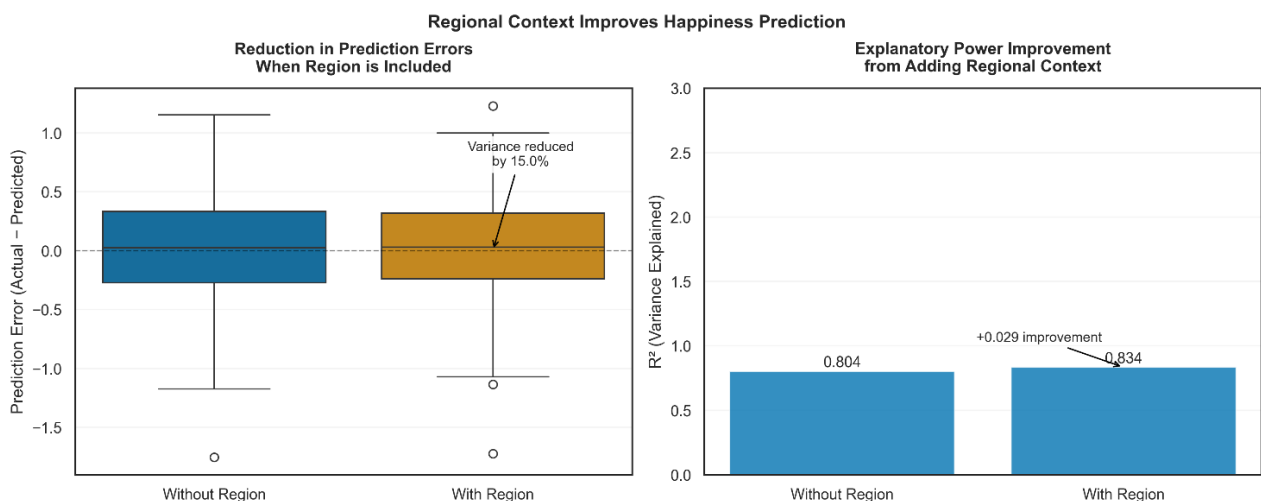


Fig. 7. Regional Clustering Prediction

This figure compares predictions from the baseline model and the model including region as an additional predictor. Incorporating region increases the model  $R^2$  from 0.804 to 0.834 and reduces

prediction error variance by 15.3%. Residuals appear more tightly grouped within regions compared to the baseline model. [11]

## RQ 2: Regional Relationship Grid

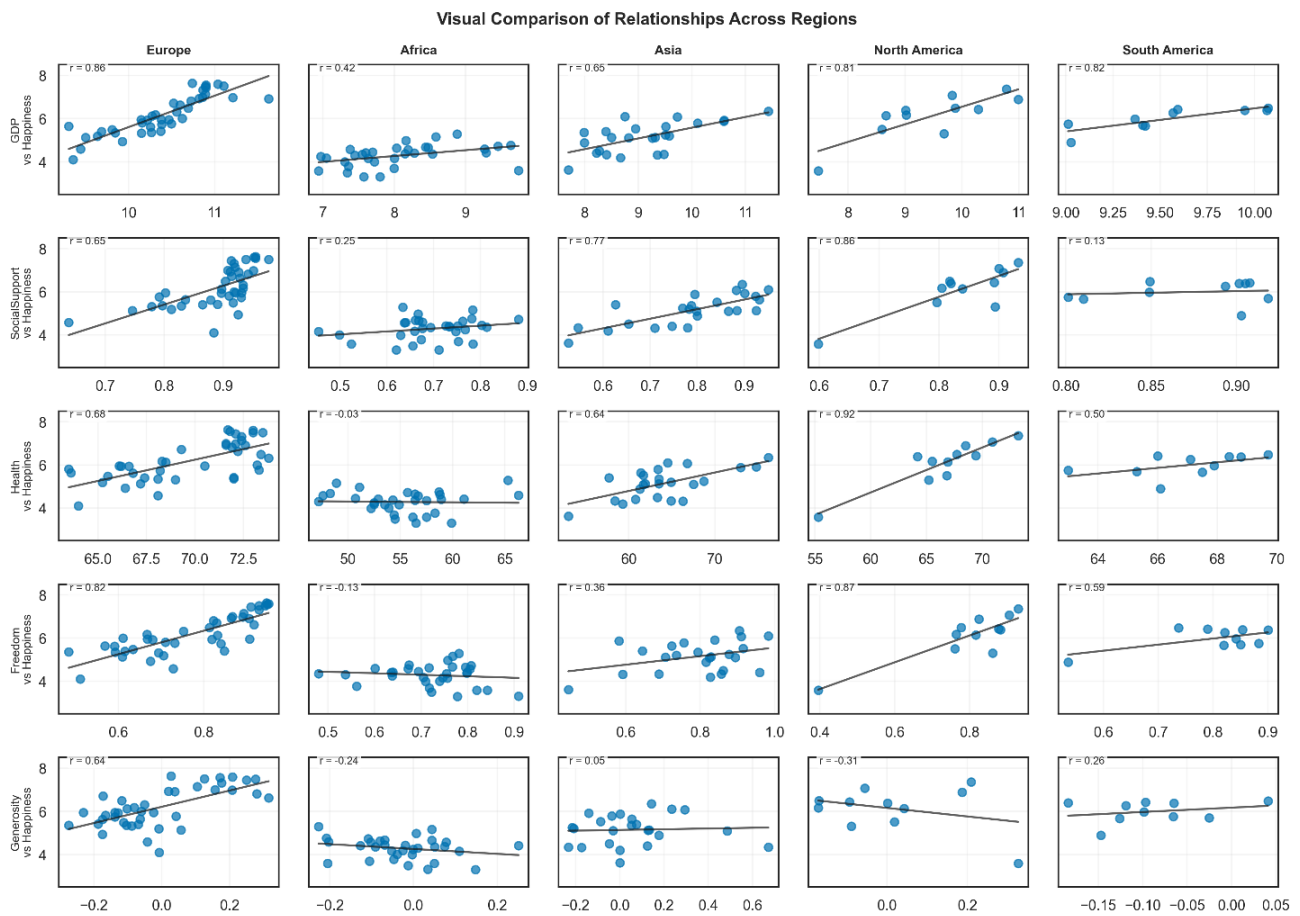


Fig. 8. Regional Variable Comparisons

The grid presents the relationships between key predictors and happiness across regions. Social support shows a positive relationship in all regions, while the slope of freedom varies across regions. Generosity displays relatively flat relationships with happiness in most regions.

RQ2: Correlation Strength Comparison

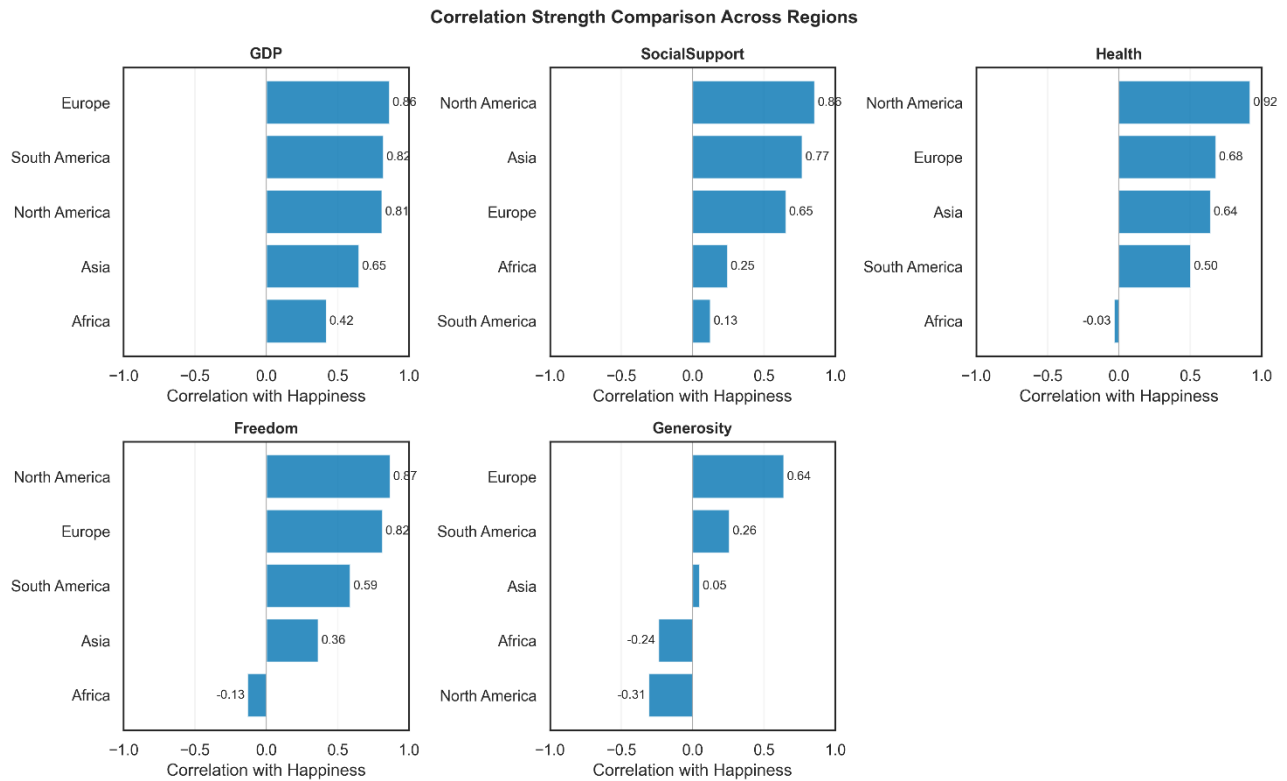


Fig. 9. Correlation Strength

Social support shows consistently high correlations, while freedom exhibits greater variation in correlation strength between regions.

RQ2: Summary Heatmap

The heatmap summarises correlation strengths between predictors and happiness by region. Social support appears as the strongest correlate in Europe, North America, Asia, and Africa, while GDP shows the strongest correlation in South America.

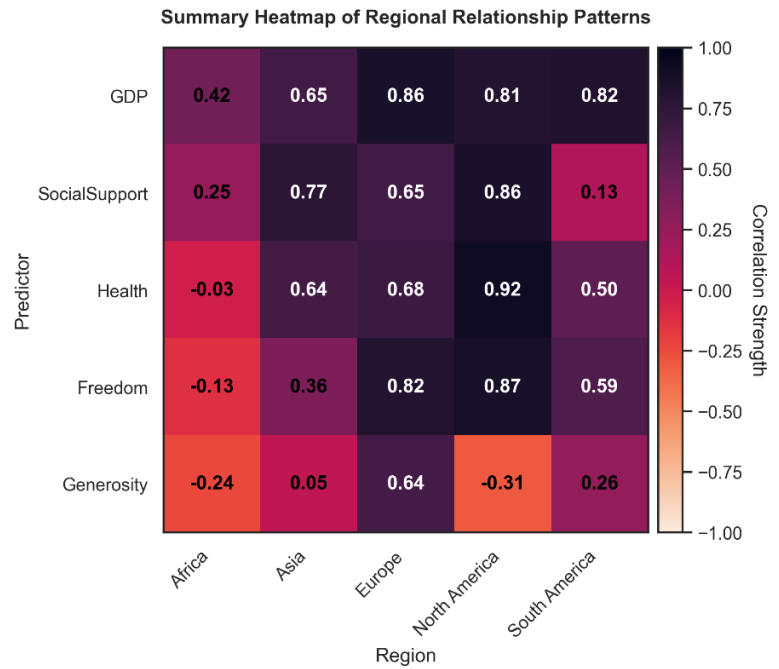


Fig. 10. Summary Heatmap

**RQ3: Happiness Relative to Wealth Predictions**

Scatterplot of actual versus GDP-predicted happiness, with countries colour-coded by performance gap. Highlights nations that over- or under-perform relative to economic expectations.

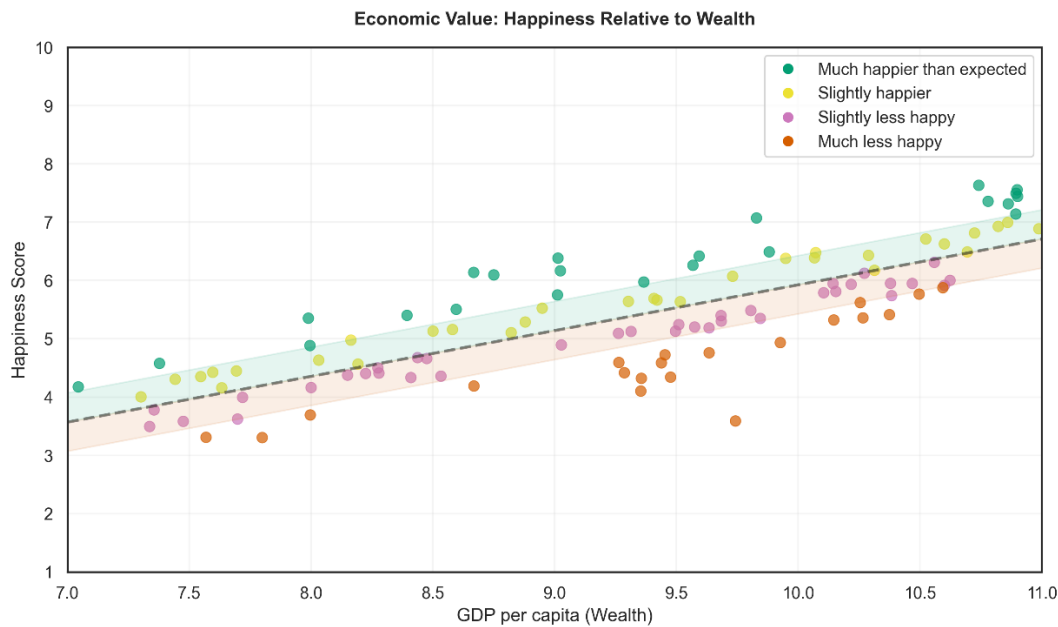


Fig. 11. Countrywide Happiness Relative to Wealth Predictions

**RQ3: Regional Economic Performance Overview**

Bubble chart aggregating countries by region. Bubble size reflects country count, while colour indicates average performance relative to wealth-based predictions.

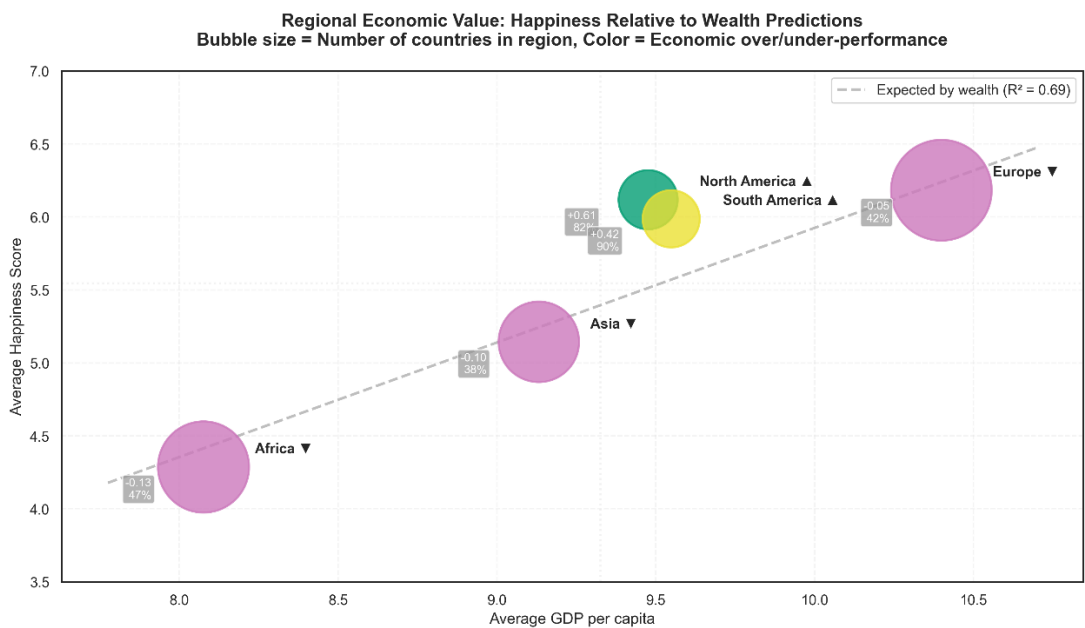


Fig. 12. Regional Happiness Relative to Wealth Predictions

### RQ3: Regional Positioning Matrix

Quadrant plot of regions by average GDP and performance gap, classifying regions by how effectively they convert economic resources into happiness. [12]

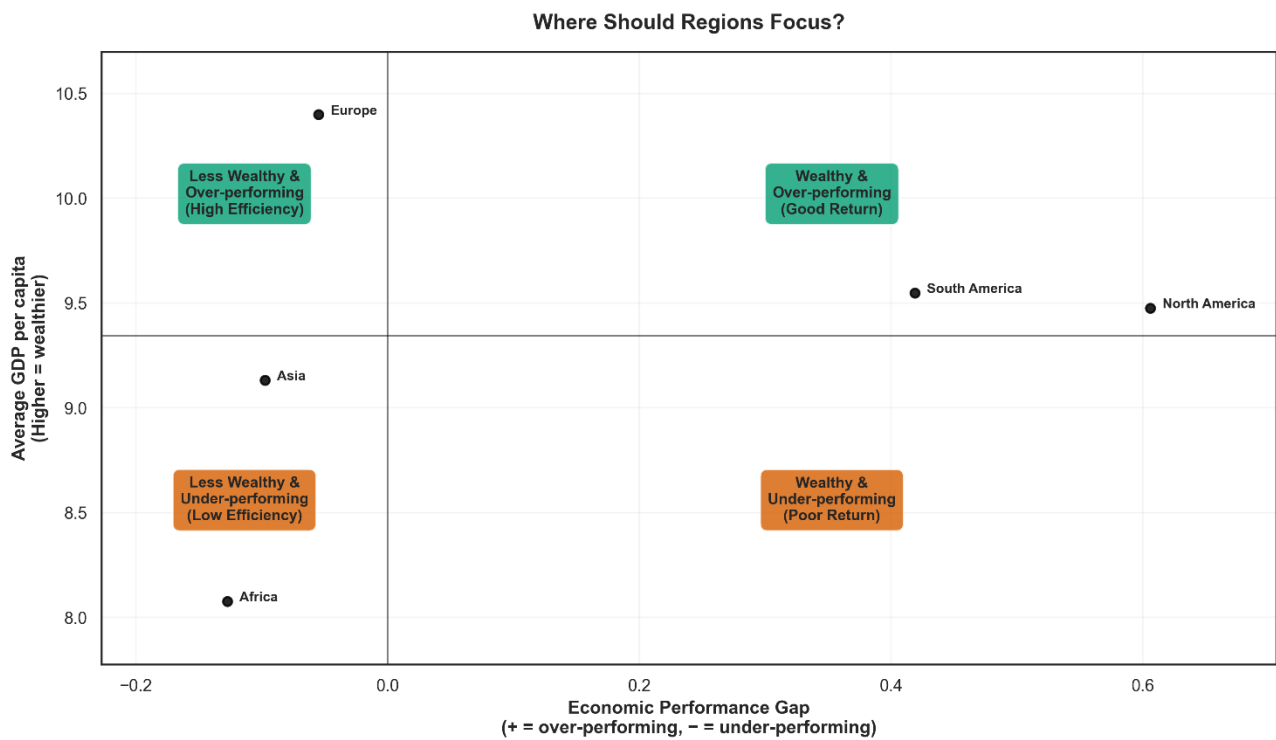


Fig. 13. Quadrant Plot of Economic Performance Gap

## 6. Discussion and Interpretation

The results reveal a clear and persistent regional pattern in national happiness that extends beyond standard economic and social explanations. When observed happiness is compared to GDP-based predictions, European countries consistently exceed expectations, while many Asian and African countries fall below them. These deviations cluster geographically rather than occurring at random, indicating that shared regional characteristics shape wellbeing outcomes. (Fig. 6)

Accounting explicitly for regional context strengthens the model substantially. The inclusion of regional indicators increases explanatory power ( $R^2$  from 0.804 to 0.834) and reduces prediction error by 15.3%, demonstrating that geography captures meaningful variation not explained by GDP, social support, health, freedom, or generosity alone (Fig. 7). This finding supports theoretical perspectives from cultural psychology and institutional economics, which argue that wellbeing is embedded in social norms, governance structures, and collective expectations. Europe's institutional emphasis on social protection and work-life balance may help explain why similar income levels yield higher life satisfaction, while rapid economic and social change in parts of Asia and Africa may dampen equivalent gains.

The analysis further shows that the mechanisms driving happiness differ across regions. Social support emerges as a consistently strong predictor worldwide, reinforcing its role as a near-universal foundation of wellbeing. In contrast, freedom exhibits pronounced regional variation, contributing strongly to happiness in Europe but weakly in Asia (Fig. 8). This divergence suggests that the meaning and value of autonomy are culturally mediated: individualistic societies appear to derive wellbeing from personal choice, whereas collectivist contexts place greater emphasis on social cohesion and role fulfillment. [13], [14] Generosity shows weak associations with happiness at the national level across all regions, implying that its effects may operate primarily at the individual scale or are insufficiently captured by aggregate indicators (Fig. 9, Fig. 10).

Evaluating happiness relative to economic expectations provides a final lens on regional performance. Europe consistently over-performs, indicating a high capacity to convert economic resources into wellbeing. Africa underperforms relative to predicted levels, likely reflecting the wellbeing costs of inequality, political instability, and limited social protection that are not captured by average GDP. Asia displays heterogeneous outcomes, with countries spanning



multiple performance categories, highlighting the region's diversity and the complex effects of modernization, rising social comparison, and cultural transition (Fig. 11, Fig. 12, Fig. 13). [15]

Overall, the findings demonstrate that happiness is not governed by universal economic relationships. Instead, it is shaped by region-specific social, cultural, and institutional contexts that influence both happiness levels and the pathways through which economic conditions translate into wellbeing. [16] This underscores the limitations of one-size-fits-all wellbeing models and highlights the importance of context-sensitive approaches to understanding and improving quality of life.

## 7. Conclusion

This study demonstrates that national happiness cannot be understood through economic indicators alone. By incorporating geographic context, the analysis moves beyond growth-centered explanations and shows that wellbeing is shaped by broader social, cultural, and institutional environments. The results support a shift away from universal models of happiness toward frameworks that recognize regional specificity.

The findings carry important implications for wellbeing research and policy. Policies aimed solely at increasing income risk overlooking the conditions that allow prosperity to translate into lived satisfaction. Instead, improving wellbeing should be focused on regional culture instead of improving the financial aspects alone as analysis also showed that social support does positively affect the happiness

Several limitations point toward future research directions. The use of national-level averages masks within-country inequality and missing or imperfect indicators may understate the role of governance and social trust. Future work could address these gaps by using subnational data, longitudinal approaches, and mixed methods to capture the lived experience of wellbeing more precisely.

Because no single metric can fully capture happiness, wellbeing measures inevitably deviate from purely objective assessment. This limitation suggests that policies to improve wellbeing should avoid universal prescriptions and instead adopt context-specific approaches tailored to each country's social, cultural, and institutional conditions.

Overall, this analysis reinforces a central insight: happiness is not simply produced by wealth, but by the ways societies organize, support, and value human life. Recognizing this opens the door to more realistic, equitable, and effective approaches to improving wellbeing worldwide. [1], [2], [15]

## References

- [1] 'CAN HAPPINESS BE MEASURED? AN OPINION PIECE', Issuu. Accessed: Jan. 06, 2026. [Online]. Available: [https://issuu.com/rgs-it/docs/welcome\\_to\\_the\\_summer\\_2023\\_edition\\_of\\_the\\_castle\\_/s/27488993](https://issuu.com/rgs-it/docs/welcome_to_the_summer_2023_edition_of_the_castle_/s/27488993)
- [2] F. Özdemir and B. Doosje, 'Happiness across cultures: Considering the relationship between general well-being and Hofstede's culture dimensions', *Int. J. Intercult. Relat.*, vol. 108, p. 102262, Sept. 2025, doi: 10.1016/j.ijintrel.2025.102262.
- [3] C. E. A. MA, 'What Is Positive and Negative Affect? Definitions + Scale', PositivePsychology.com. Accessed: Jan. 06, 2026. [Online]. Available: <https://positivepsychology.com/positive-negative-affect/>
- [4] 'What is the difference between linear and nonlinear interpolation? Can linear interpolation be used for nonlinear data? If so, how can it...?', Quora. Accessed: Jan. 06, 2026. [Online]. Available: <https://www.quora.com/What-is-the-difference-between-linear-and-nonlinear-interpolation-Can-linear-interpolation-be-used-for-nonlinear-data-If-so-how-can-it-be-done-efficiently>
- [5] 'Interpolation (scipy.interpolate) — SciPy v1.16.2 Manual'. Accessed: Jan. 06, 2026. [Online]. Available: <https://docs.scipy.org/doc/scipy/reference/interpolate.html>
- [6] 'What is Exploratory Data Analysis?', GeeksforGeeks. Accessed: Jan. 06, 2026. [Online]. Available: <https://www.geeksforgeeks.org/data-analysis/what-is-exploratory-data-analysis/>
- [7] 'What is Exploratory Data Analysis? | IBM'. Accessed: Jan. 06, 2026. [Online]. Available: <https://www.ibm.com/think/topics/exploratory-data-analysis>
- [8] 'Colorblind Safe Color Schemes'.
- [9] 'Finding the Best Color-blind Friendly Palette on Python Seaborn | by Boriharn K | Data Science Collective | Medium'. Accessed: Jan. 06, 2026. [Online]. Available: <https://medium.com/data-science-collective/finding-the-best-color-blind-friendly-palette-on-python-seaborn-0546e4ed33f3>
- [10] 'Choosing the Correct Type of Regression Analysis - Statistics By Jim'. Accessed: Jan. 06, 2026. [Online]. Available: <https://statisticsbyjim.com/regression/choosing-regression-analysis/>
- [11] Atlassian, 'A Complete Guide to Box Plots', Atlassian. Accessed: Jan. 06, 2026. [Online]. Available: <https://www.atlassian.com/data/charts/box-plot-complete-guide>
- [12] B. Q. Ford *et al.*, 'Culture shapes whether the pursuit of happiness predicts higher or lower well-being', *J. Exp. Psychol. Gen.*, vol. 144, no. 6, pp. 1053–1062, Dec. 2015, doi: 10.1037/xge0000108.
- [13] J. Jiao and J. Zhao, 'Individualism, Collectivism, and Allocation Behavior: Evidence from the Ultimatum Game and Dictator Game', *Behav. Sci.*, vol. 13, no. 2, p. 169, Feb. 2023, doi: 10.3390/bs13020169.
- [14] that1gingergirl, 'Individualism vs. collectivism is a important concept in Social Psychology. It effects every facet of a culture, including how individuals view themselves and the world. Where does your world fall on this spectrum?', r/worldbuilding. Accessed: Jan. 06, 2026. [Online]. Available: [https://www.reddit.com/r/worldbuilding/comments/q0f5s7/individualism\\_vs\\_collectivism\\_is\\_a\\_important/](https://www.reddit.com/r/worldbuilding/comments/q0f5s7/individualism_vs_collectivism_is_a_important/)
- [15] H. Choi and E. Choi, 'Unraveling Why Happiness Levels Vary Across Cultures: Mechanisms Underlying East-West Differences', *Soc. Personal. Psychol. Compass*, vol. 19, no. 9, p. e70078, 2025, doi: 10.1111/spc3.70078.
- [16] 'Is happiness relative? Can happiness be measured objectively? | Notes and Queries | guardian.co.uk'. Accessed: Jan. 06, 2026. [Online]. Available: <https://www.theguardian.com/notesandqueries/query/0,5753,-26803,00.html>

[17] OpenAI, *ChatGPT* [Large language model]. [Online]. Available: <https://openai.com/chatgpt>

[18] DeepSeek, AI research assistant software. [Online]. Available: <https://www.deepseek.com>

*Note: ChatGPT and DeepSeek were used solely to support brainstorming, idea generation, and feedback on structure and clarity. No AI-generated content was included in the coursework without human review.*

# Appendices

## Appendix A: Data Preparation Outputs

This appendix documents the steps and files involved in preparing the dataset for analysis.

**File A1: Raw Dataset:** [df\\_raw.html](#)

The initial dataset collected for analysis, before any cleaning or preprocessing.

**File A2: Standardized variable names:** [df\\_prep.html](#)

Standardized variable names for readability

**File A3: Raw Missing Data Before Interpolation:** [missing\\_raw.html](#)

Overview of missing values across variables prior to any imputation.

**File A4: Missing Data After Interpolation:** [missing\\_interpolated.html](#)

Dataset after interpolation, showing how missing values were addressed.

**File A5: Excluded Countries:** [excluded\\_countries.html](#)

List of countries removed during preprocessing due to incomplete or insufficient data.

**File A6: Final Dataset:** [preprocessed\\_dataset.html](#)

The cleaned and filtered dataset, ready for analysis.

## Appendix B: Colourblind-Friendly Guide

Files that involved in designing colourblind-friendly visualization

**File B1: PDF of palettes:** [Colorblind Safe Color Schemes.pdf](#)

Colourblind Schemes by NCEAS Science Communication Resource Corner, Alexandra Phillips

## Appendix C: GitHub Repo for Codebase

Link of the Repository created for artefact of this report

**Link C1:** [GitHub Repo](#)