DATA ANALYSIS CASE STUDY

Happiness, Socioeconomic Conditions and Human Outcomes

An exploratory analysis of international wellbeing data

Analyst Reis Tyson

Tools Python, pandas, NumPy, SciPy, Matplotlib

Methods Statistical modelling, regression, model comparison, subgroup analysis

About this analysis.
This short case study explores relationships between national happiness, economic conditions, social support and life expectancy.
It was originally completed in under one week as part of a university statistical data analysis module, alongside my MPhys research and other taught work. The statistical methods are intentionally transparent and interpretable: the aim was not to maximise model complexity, but to understand what conclusions the available data could reasonably support.
The analysis was conducted independently in Python. This report presents the principal findings in accessible language while retaining the underlying statistical results for technical scrutiny.

What seems to influence national happiness?

The dataset contains country-level measures of happiness alongside economic and social indicators, including logged GDP per capita, social support and life expectancy.

Across the sample, the average happiness score was

$$\mathbf{5.544},$$

with a median of

$$\mathbf{5.694}.$$

I began with one of the most intuitive relationships in the data:

Are people in wealthier countries generally happier?

Wealth and happiness are positively associated

A clear positive relationship was present between logged GDP per capita and national happiness.

A weighted linear model gave a positive gradient of

$$0.910 \pm 0.012.$$

Higher GDP per capita was therefore generally associated with higher reported happiness.

The fitting procedure incorporated the reported uncertainty in the happiness measurements rather than assuming that every observation was equally precise.

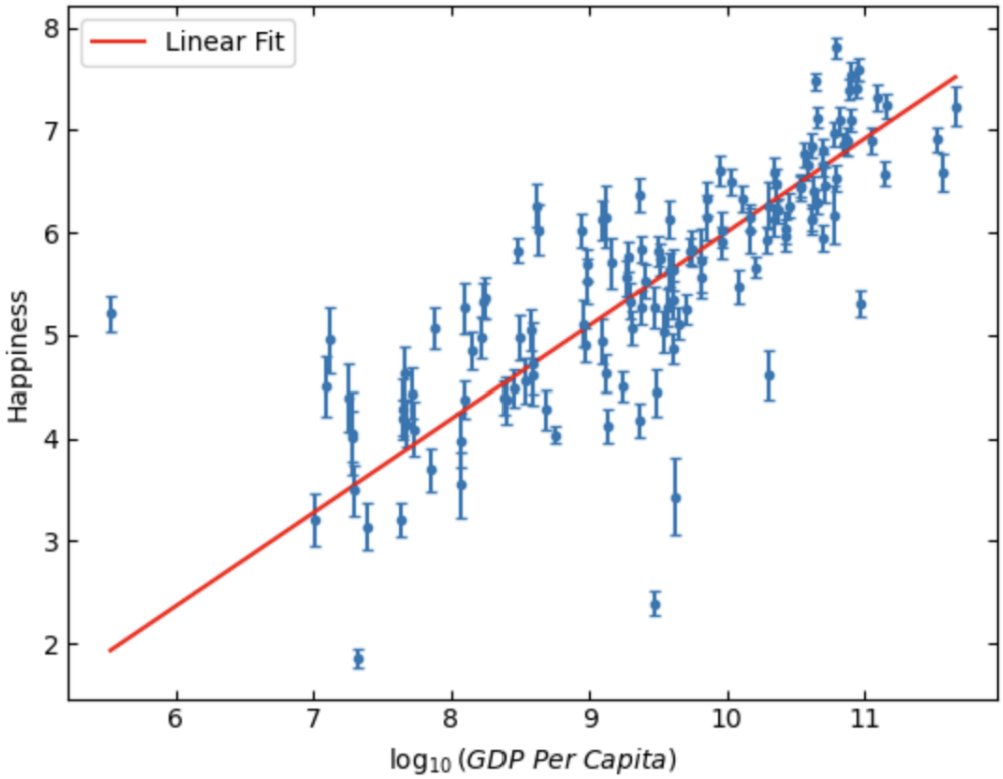

Figure 1: National happiness as a function of logged GDP per capita. The relationship is clearly positive, but substantial variation remains between countries.

Key finding: Wealth is clearly associated with happiness, but GDP alone does not explain the substantial differences between countries.

This raised a second question: could a more flexible mathematical model describe the relationship better?

Does a more complex model tell us more?

I compared four models of increasing complexity: linear, exponential, quadratic and cubic.

Their performance was assessed using reduced chi-squared (χ^2_{red}), which measures how closely a model describes the observations relative to their reported uncertainties.

Table 1: Comparison of models fitted to happiness as a function of logged GDP per capita.

Model	χ^2	χ^2_{red}
Linear	3138.62	23.42
Exponential	3018.43	22.53
Quadratic	2983.56	22.43
Cubic	2848.53	21.58

Numerically, the cubic model performed best.

But that did not necessarily make it the most credible model.

The best statistical score was not necessarily the best explanation

Inspecting the fitted models revealed that the additional flexibility of the cubic model allowed observations in the sparsely sampled lowest-GDP region to have a disproportionate influence on its shape.

The resulting behaviour could represent a genuine feature of the population. However, there were not enough observations in this region to establish that confidently.

I therefore did **not** interpret the lowest reduced chi-squared value as sufficient evidence to prefer the cubic model.

> **Analytical lesson:**
> Model selection requires both quantitative testing and critical inspection of the data. A more complicated model can produce a better numerical fit without necessarily providing a more reliable explanation.

This is particularly important when working with human data, where sampling and population structure can influence apparently strong statistical patterns.

Is social support more closely related to happiness?

I next examined reported social support.

Again, the relationship was clearly positive: countries reporting greater social support tended to report greater happiness.

A weighted linear fit produced a reduced chi-squared of

$$\chi^2_{\text{red}} = \mathbf{14.32},$$

compared with

$$\chi^2_{\text{red}} = \mathbf{23.42}$$

for the equivalent linear relationship between GDP and happiness.

Under this simple modelling framework, social support therefore described the observed happiness trend more closely than GDP alone.

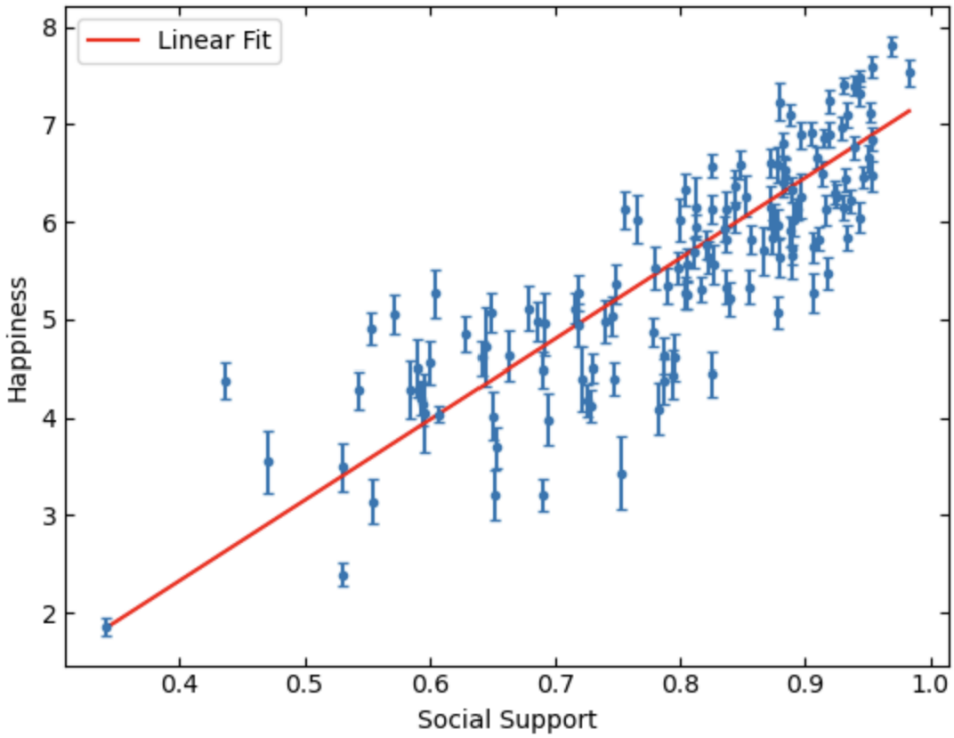

Figure 2: Relationship between reported social support and national happiness. A strong positive trend is visible, although considerable variation remains unexplained by a single predictor.

However, neither relationship provided a complete explanation of happiness. Considerable variation remained between countries.

> **Key finding:**
> Neither wealth nor social support should be interpreted as a single explanation for national happiness. Human outcomes emerge from multiple, interacting factors.

This led to the more interesting extension of the analysis.

Is social support associated with longer lives — or are we simply measuring wealth?

Countries with greater social support also appeared to have greater life expectancy.

On its own, however, this relationship is difficult to interpret.

Wealthier countries may have better healthcare, infrastructure, nutrition and other advantages. If wealthy countries also report greater social support, an apparent relationship between social support and life expectancy could partly reflect differences in wealth.

I therefore tested this before interpreting the relationship.

Wealth is strongly related to both variables

Logged GDP per capita was strongly associated with life expectancy:

$$R^2 = 0.701, \qquad p = 5.47 \times 10^{-37}.$$

In accessible terms, approximately **70% of the observed variation** in life expectancy was associated with variation in logged GDP within this simple relationship.

GDP was also associated with social support:

$$R^2 = 0.550, \qquad p = 5.20 \times 10^{-25}.$$

Wealthier countries therefore tended, on average, to have **both longer life expectancy and greater reported social support**.

This meant that comparing all countries directly could give a misleading impression of the relationship between the two.

> **The analytical problem:**
> If wealth is related to both social support and life expectancy, how can we investigate the relationship between social support and life expectancy without simply rediscovering the difference between richer and poorer countries?

Making the comparison fairer

Rather than directly comparing countries with very different economic circumstances, I divided the sample into two equally sized groups:

- **lower-GDP countries**; and
- **higher-GDP countries**.

Within each group, social support and life expectancy were measured relative to the typical value for countries in that economic group.

Technically, each variable was normalised by its group mean:

$$S_{\text{relative}} = \frac{S}{\bar{S}_{\text{group}}},$$

and

$$L_{\text{relative}} = \frac{L}{\bar{L}_{\text{group}}}.$$

The question therefore changed from:

> *Do countries with greater social support have greater life expectancy?*

to the more useful:

> **Among countries with broadly similar economic circumstances, is relatively greater social support still associated with relatively greater life expectancy?**

This does not remove every possible confounding factor, but it reduces the extent to which the result simply compares richer countries with poorer ones.

Linear regressions were then performed independently for the two groups.

The relationship remains within both economic groups

A statistically significant positive relationship remained in both subsets.

Lower-GDP countries

26.5%

of the variation in relative life expectancy was associated with relative social support within this simple linear model.

$$R^2 = 0.265, \qquad p = 6.94 \times 10^{-6},$$

with a fitted slope of

$$0.222 \pm 0.045.$$

Higher-GDP countries

14.9%

of the variation in relative life expectancy was associated with relative social support.

$$R^2 = 0.149, \qquad p = 1.13 \times 10^{-3},$$

with a fitted slope of

$$0.293 \pm 0.086.$$

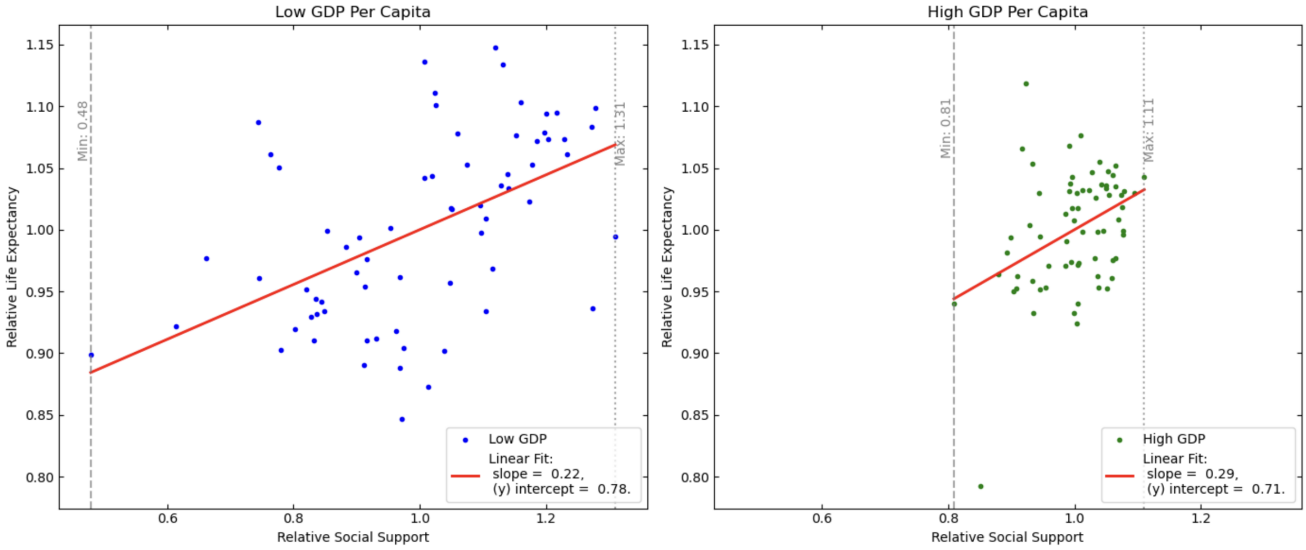

Figure 3: Relative social support against relative life expectancy for lower- and higher-GDP countries. A positive relationship remains within both economic groups.

Key finding:
The positive relationship between social support and life expectancy persists when lower- and higher-GDP countries are considered separately.
The overall association therefore cannot be attributed solely to comparing wealthy countries with poorer ones.

There is, however, an important qualification.

Why the two groups should not be compared too quickly

At first glance, the lower-GDP group appears to show the stronger relationship because its R^2 value is higher.

But the two datasets do not cover the same range.

Relative social support in the higher-GDP countries occupied a relatively narrow range of approximately

$$0.81 S_{\text{relative}} 1.11,$$

while the lower-GDP countries extended approximately from

$$0.48 S_{\text{relative}} 1.31.$$

This restricted range gives the higher-GDP sample less statistical leverage with which to characterise a relationship.

The fitted slopes were

$$\text{Lower GDP: } 0.222 \pm 0.045$$

and

$$\text{Higher GDP: } 0.293 \pm 0.086.$$

Their uncertainties overlap.

I therefore would not claim from this analysis that social support has a meaningfully different relationship with life expectancy in richer and poorer countries.

A defensible interpretation:
A positive association between relative social support and relative life expectancy is present within both economic groups, but the available data do not establish a causal relationship or a clear difference between the two populations.

What can we conclude?

This exploratory analysis identified several patterns.

Greater national wealth was associated with greater reported happiness, while social support also showed a substantial positive relationship with happiness.

The extension demonstrated why apparent relationships in human data need to be treated carefully. Social support and life expectancy initially appeared strongly related, but both were also associated with national wealth.

Separating countries into broad economic groups provided a more informative comparison. A positive association between relative social support and relative life expectancy remained in both groups, suggesting that the global relationship is not solely a consequence of differences in GDP.

The analysis does **not** demonstrate that social support causes increased life expectancy. Other socioeconomic, demographic and institutional variables could influence the observed relationships.

A larger follow-up study could move beyond these one- and two-variable comparisons using multivariate statistical modelling and, where appropriate, more advanced predictive methods.

An interesting difference between the two groups

There is a further feature of the results that warrants investigation.

The fitted relationship between relative social support and relative life expectancy was

$$m_{\text{lower GDP}} = 0.222 \pm 0.045$$

compared with

$$m_{\text{higher GDP}} = 0.293 \pm 0.086.$$

The point estimate is therefore approximately 32% steeper in the higher-GDP group.

This is particularly interesting because the higher-GDP countries occupy a much narrower range of relative social support:

$$0.81 S_{\text{relative}} 1.11,$$

compared with approximately

$$0.48 S_{\text{relative}} 1.31$$

among lower-GDP countries.

The higher-GDP group consequently has a lower R^2, meaning that the simple linear relationship accounts for a smaller proportion of the observed variation. However, this should not be confused with the magnitude of the fitted slope: within the observed range, the estimated relationship is actually steeper in the wealthier group.

> **An exploratory finding, not yet a conclusion:**
> The point estimates suggest that relative social support may have a stronger association with life expectancy among higher-GDP countries.
> However, the uncertainties on the two fitted slopes overlap substantially. This analysis therefore does not provide sufficient evidence to conclude that the underlying relationships are genuinely different.

A hypothesis for further investigation

One possible explanation is that the relative importance of social and material conditions changes with national wealth.

In lower-GDP countries, constraints on material conditions necessary for healthy survival — potentially including healthcare, nutrition, infrastructure and sanitation — may place a stronger limit on life expectancy. Greater social support may therefore be unable to compensate fully for insufficient material resources.

Among wealthier countries, a greater proportion of the population may already have access to the material conditions required for healthy survival. Once these requirements are more consistently met, differences in social conditions could potentially become relatively more important in explaining differences in health outcomes.

This interpretation is a hypothesis generated by the exploratory results, rather than a conclusion supported by the present analysis. Testing it would require a larger study with explicit socioeconomic controls and a formal test of whether the relationship between social support and life expectancy changes with GDP.

> **Next analytical question:**
> Does national wealth moderate the relationship between social support and life expectancy?
> A follow-up analysis could test this directly using a multivariate model containing GDP, social support and an interaction between the two, rather than dividing countries into two groups.

Analytical approach

This study demonstrates how I approach an unfamiliar dataset.

Explore the data
Identify distributions, relationships and potential problems before modelling.

Use uncertainty where available
The fitting procedure incorporated reported measurement uncertainty rather than treating all observations as equally precise.

Test rather than assume
Multiple candidate models were quantitatively compared.

Inspect the model as well as its score
The numerically best-performing model was not automatically accepted when its behaviour was poorly constrained by the available data.

Look for confounding structure
The relationship between GDP and both variables was investigated before interpreting social support against life expectancy.

Adapt the analysis to the problem
The population was divided into economic groups when the initial comparison proved difficult to interpret.

Communicate uncertainty
Statistical association was not presented as evidence of causation.

Technical note

The analysis was implemented independently in **Python** using **pandas, NumPy, SciPy and Matplotlib**.

Methods included data manipulation, exploratory analysis, uncertainty-weighted nonlinear curve fitting, custom chi-squared evaluation, linear regression, model comparison, normalisation, subgroup analysis and statistical visualisation.

The intentionally interpretable methods used here were appropriate to the scope and size of the study. My broader analytical experience extends to higher-dimensional statistical techniques and automated analysis pipelines, and this workflow can be developed toward multivariate or machine-learning methods where the data and research question justify the additional complexity.

Code availability. The complete Python analysis notebook can be provided on request.

This case study is an edited presentation of exploratory analysis originally completed as university coursework. It is presented as an example of analytical workflow, statistical reasoning and communication rather than as a definitive socioeconomic study.