

Final Project

Chris Orellana (c_orellana@ucsb.edu)
Andrew Guerra (andrewguerra@umail.ucsb.edu)
Austin Zhang (xinhaozhang@ucsb.edu)
Rachel Li (ruiqi_li@umail.ucsb.edu)

Abstract

“How does the relationship between key predictors and happiness vary across world regions, and what might this imply about region-specific drivers of wellbeing?”

In this report, we explore how the strength of various predictors of happiness varies across global regions. Using data from the 2023 World Happiness Report, which includes yearly observations from over 160 countries between 2008 and 2023, we analyze both regional patterns and global trends in the determinants of subjective wellbeing. Our approach combines exploratory data visualization with regression modeling to assess whether certain factors have a stronger association with happiness in some regions than in others.

Introduction

The core of the World Happiness Report (WHR) is a metric known as the Cantril Self-Anchoring Scale, in which respondents are asked:

“Please imagine a ladder with steps numbered from 0 at the bottom and 10 at the top. The top of the ladder represents the best possible life for you and the bottom of the ladder represents the worst possible life for you. On which step of the ladder would you say you personally feel you stand at this time?”

This question avoids explicit references to “happiness” or “well-being”, making it broadly interpretable and culturally adaptable. A score of 10 represents the best possible life, while 0 represents the worst. By asking a broad question that ranks from best to worst, we preserve simplicity which makes it suitable for cross-cultural comparison at a global scale. After individual scores are collected for the country, the results are averaged to provide a score representing the country’s overall happiness ranking. This score is stored as the **Life Ladder variable**.

In addition to the Cantril Ladder scores, the dataset includes 8 variables that serve as predictors of well-being: **Log GDP per Capita**, **Social support**, **Healthy life expectancy at birth**, **Freedom to make life choices**, **Generosity**, **Perceptions of corruption**, **Positive affect**, and **Negative affect**. Together, these variables provide insight into the conditions that shape life satisfaction.

This project investigates whether the influence of these predictors is consistent across regions. We aim to identify whether certain variables are more or less predictive of happiness depending on a region's social and cultural context.

Explorations & Analysis

To understand the factors associated with global happiness, we began our analysis by examining the size of the data set and the distributions of **key predictors** included in the World Happiness Report. By exploring the predictors' distributions, we aim to assess the range, variation, and potential skewness of the data. These initial observations provide valuable context for understanding how these factors may relate to self-reported happiness.

We first created a table to display the number of rows, years, and countries included in the data set.

total_rows	total_years	total_countries
2199	18	165

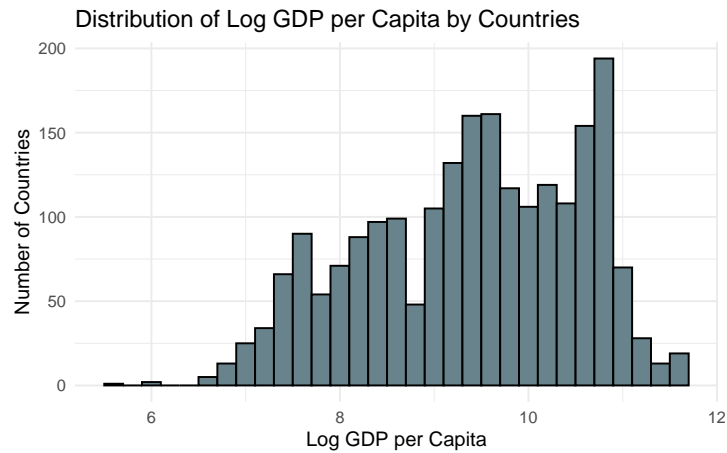
With 165 different countries represented, we have data spanning across 18 years (2005-2022) that contain 2,199 rows of information regarding our 8 different variables.

As previously stated, our Life Ladder variable is the Happiness level. A series of 8 different predictors are used in association with this value that are as follows.

COMMENTS ON DIST - GDP

- **Log GDP Per Capita:** This predictor variable represents a country's Gross Domestic Progress (GDP), which measures the average economic output per person and serves as a rough indicator of a country's standard of living.

A log-transformation has been applied to help reduce skewness. This transformation makes it easier to observe more sensitive differences in GDP Per Capita across countries.

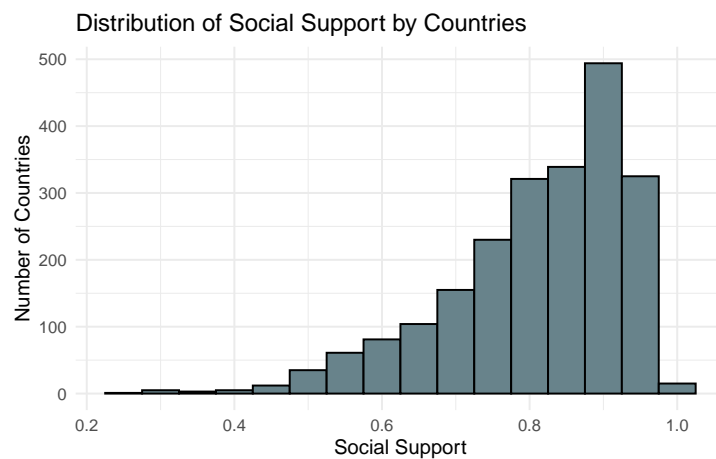


HISTOGRAM ANALYSIS

With the majority of nations grouped between 8 and 10, the log GDP per capita histogram is somewhat skewed to the left, suggesting that moderate economic output per person is typical around the world. A lesser number of nations, probably from affluent regions like Western Europe and North America & ANZ, fall outside of this range. Our research question is supported by this distribution, which demonstrates that although GDP is a good indicator of happiness, its unequal distribution across regions emphasizes the need to take other factors into account, particularly in lower-income areas where other factors like social support may have a greater impact on wellbeing.

COMMENTS ON DIST - SOCIAL SUPPORT

- **Social Support:** This variable captures the perceived availability of reliable social connections. Higher values indicate a stronger sense of social support, which is often associated with greater emotional well-being and resilience in the face of hardship.



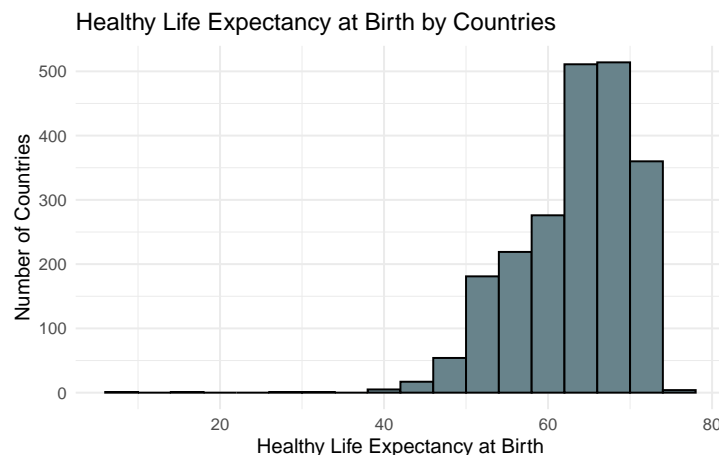
HISTOGRAM ANALYSIS

The majority of nations report high levels (over 0.75), and the social support histogram displays a moderate left-skewed distribution. This implies that social networks are typically seen as robust by individuals all across the world. According to the World Happiness Report, social support is a major factor in wellbeing, particularly in areas with lower GDPs. The distribution lends credence to the notion that global happiness is significantly influenced by relational and emotional elements in addition to economic ones.

COMMENTS ON DIST - LIFE EXPECTANCY

- **Healthy Life Expectancy at Birth:** This variable represents the average number of years a person born in a given country can expect to live in good health. It accounts the total length of life as well as quality of those years, adjusting for years lived in deteriorated health due to disease or disability.

Higher values indicate better overall health conditions and access to healthcare, and it serves as a key indicator of national well-being and public health infrastructure.



HISTOGRAM ANALYSIS

The histogram is left-skewed, with most countries concentrated between 60–70 years, reflecting generally good worldwide health. The long tail on the lower end, however, reveals notable differences, indicating that healthcare quality and access continue to differ greatly and can affect the well-being of a country.

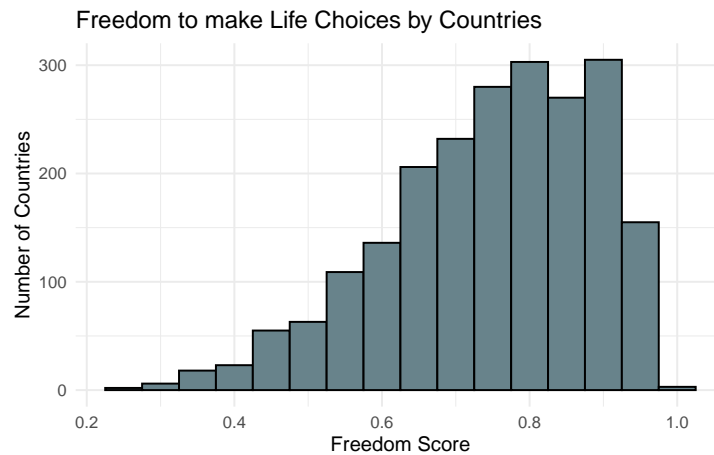
COMMENTS ON DIST - FREEDOM OF LIFE CHOICES

- **Freedom to make Life Choices:** This variable is a 0 to 1 measure of a country's average response to the question:

Are you satisfied or dissatisfied with your freedom to choose what you do with your life?

Higher values indicate that more people in a country feel they have greater personal autonomy and decision-making.

This variable reflects the perceived ability of individuals to shape their own lives, which is closely linked to overall life satisfaction.



HISTOGRAM ANALYSIS

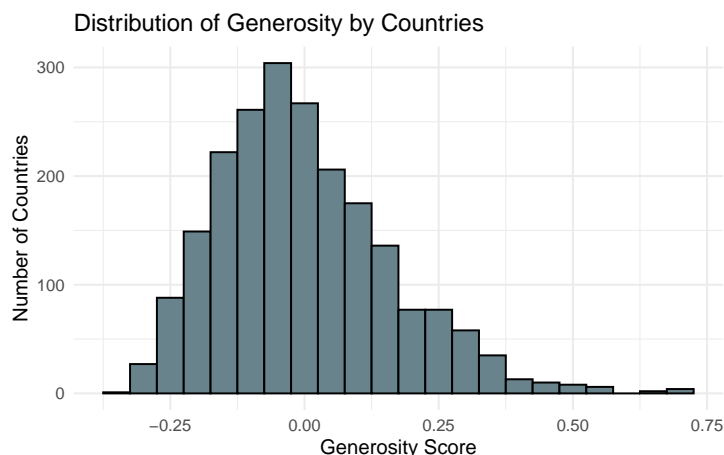
With a clear bias toward scores above 0.6, the histogram demonstrates that the majority of nations report high levels of felt freedom to make life decisions. This implies that people believe they have a great deal of personal autonomy in many different places. The disparity at the lower end, however, suggests that this autonomy may be restricted in certain nations—likely those with less political or social freedom. In relation to our research issue, this lends credence to the notion that the impact of freedom on pleasure varies and might be especially significant in areas with a deficiency of other structural supports.

COMMENTS ON DIST - GENEROSITY

- **Generosity:** This variable reflects the extent to which people in a country report recent acts of generosity, such as donating to charity.

However, rather than using raw percentages, the World Happiness Report calculates generosity as the residual from a regression that adjusts for GDP Per Capita. This

means values can be positive or negative, depending on whether a country is more or less generous than expected based on its income. Higher values indicate greater-than-expected generosity, reflecting strong pro-social behavior and social capital.



HISTOGRAM ANALYSIS

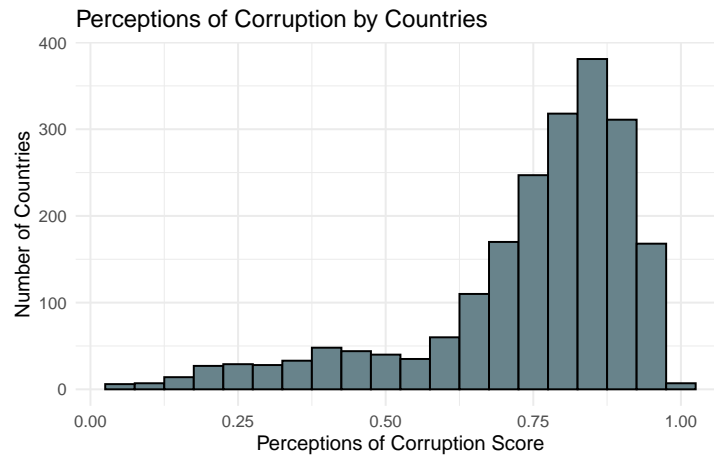
Most nations cluster around a generosity score close to zero, indicating a slightly left-skewed generosity histogram. This implies that the generosity of the majority of nations is about what one would anticipate given their wealth levels. Fewer nations have notably positive ratings, which indicate generosity that is above expectations, while a few have negative scores, which indicate generosity that is below expectations. This distribution illustrates the differences in social capital and pro-social conduct between countries despite comparable economic circumstances, which is consistent with the notion of generosity as a residual.

COMMENTS ON DIST - CORRUPTION

- **Perceptions of Corruption:** The average of two questions:

“Is corruption widespread throughout the government or not?” and “Is corruption widespread within businesses or not?”

This variable is used as the overall corruption-perception measure.

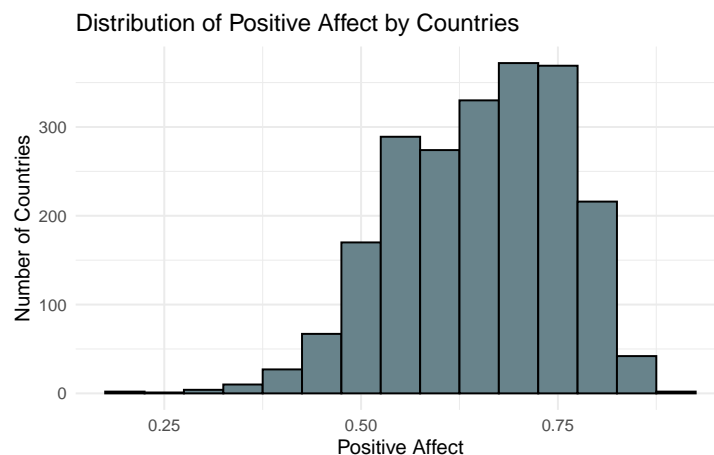


HISTOGRAM ANALYSIS

The majority of nations scored higher than 0.6 on the perceptions of corruption histogram, which displays a distribution that is skewed to the right. This implies that most nations believe corruption to be at a relatively low level. Strong public trust in institutions is shown in many places by the peak, which falls between 0.75 and 0.85. This supports the report's conclusion that higher scores indicate a better level of trust in the honesty of the public and private sectors.

COMMENTS ON DIST - POSITIVE AFFECT

- **Positive Affect:** The national average of binary responses (0=no, 1=yes) about three emotions experienced on the previous day: **laughter**, **enjoyment**, and **interest**.

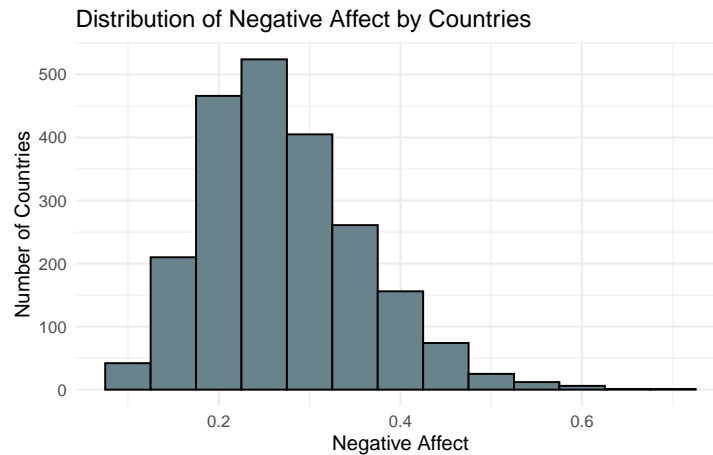


HISTOGRAM ANALYSIS

Most nations fell between 0.55 and 0.75 on the happy Affect histogram, indicating a high degree of reported happy feelings overall. We may imply that the effect of positive effect is constant across different regions.

COMMENTS ON DIST - NEGATIVE AFFECT

- **Negative Affect:** The national average of binary responses (0=no, 1=yes) about three emotions experienced on the previous day: **worry, sadness, and anger**.

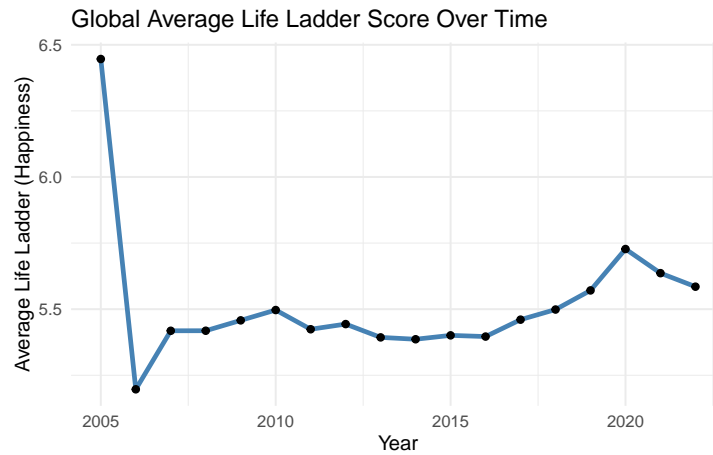


HISTOGRAM ANALYSIS

The majority of nations have comparatively low levels of unpleasant emotional experiences, according to the negative affect histogram, which has a right-skewed distribution. The peak falls between 0.2 and 0.3, indicating that people in many nations report low frequency of unpleasant emotions like melancholy, anger, or concern. Extreme emotional anguish is rare in the dataset, as only a small number of nations report significant levels of negative affect.

LIFE LADDER OVER TIME

To explore how global happiness has changed over time, we examined the distribution of Life Ladder scores across the years included in the data set. To do so, we calculated the grand mean of each country and grouped the values by year and generated a line plot to observe year-to-year trends.



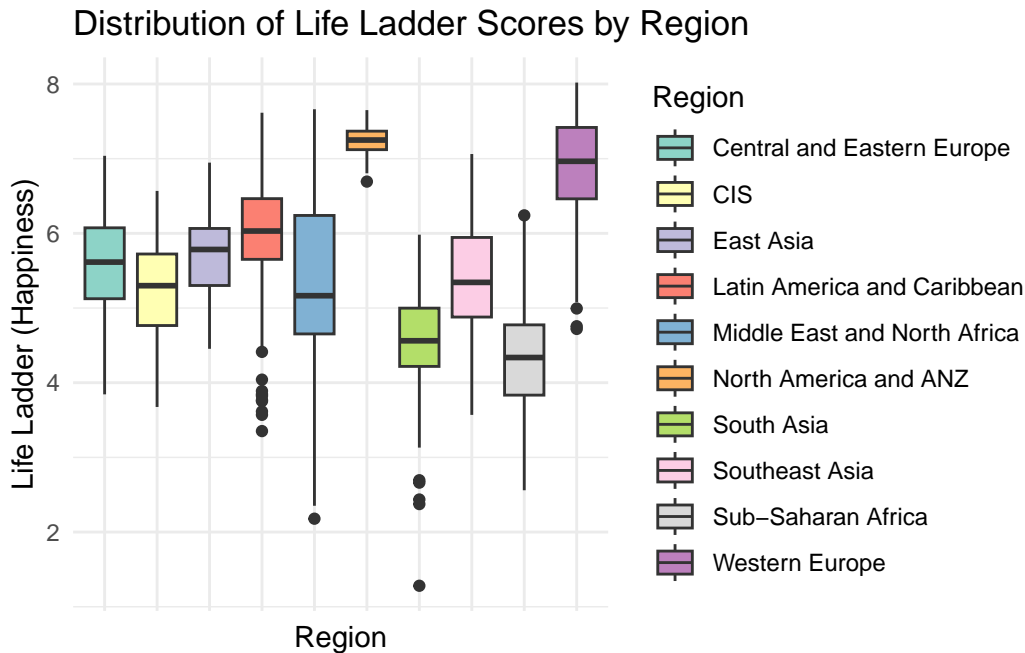
Based on our Line Plot, we see that global happiness saw slightly over a 1.0 drop from 2005 to 2006, which may have been due to external factors relating to the sampling of data and improved methodology rather than a true shift in well-being. After 2007, happiness levels have remained relatively stable with slight upward trends up until 2020. After 2020, happiness levels saw a gradual decrease up until 2022 which reflects world-wide disruptions resulting from the COVID-19 pandemic. Overall, trends appear very minimal.

After observing the distribution of all variables as well as life ladder ratings across the years, we now separated our data into world regions, as designated by the World Happiness Report.

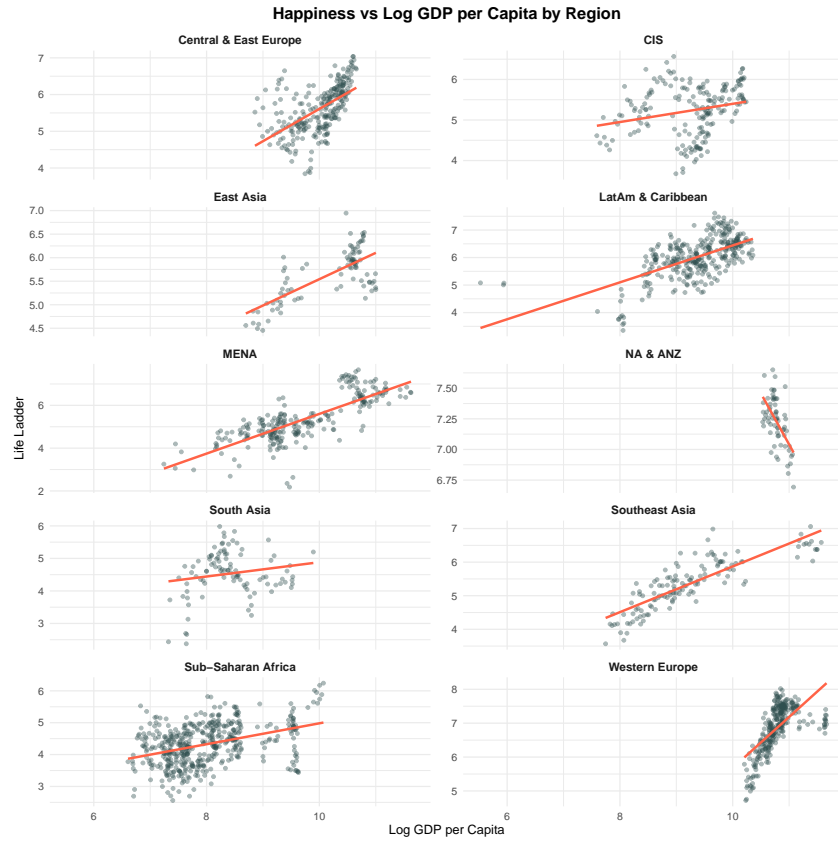
x

South Asia
 Central and Eastern Europe
 Middle East and North Africa
 Sub-Saharan Africa
 Latin America and Caribbean
 CIS
 North America and ANZ
 Western Europe
 Southeast Asia
 East Asia

The World Happiness Report uses the 10 regions described above. With these regions, we aim to see if any variables hold varied significance across the world. To begin with, we visualized the distribution of Life Ladder scores by Region.



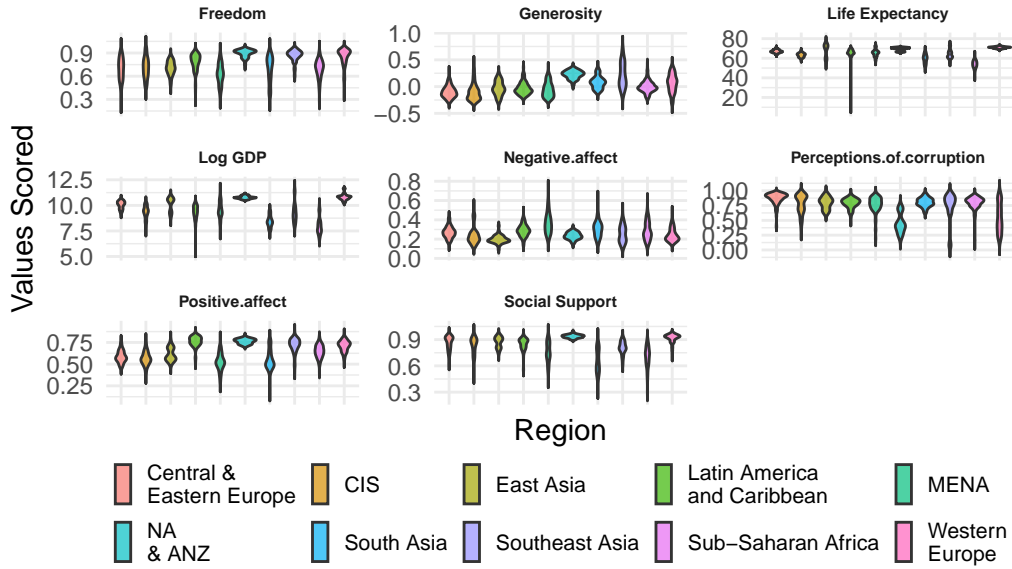
Based on our side-by-side Boxplot of all 10 regions, we see that Western Europe as well as North America and ANZ average significantly higher happiness scores than their non-western counterparts. While it could reflect a difference in happiness score from western and non-western countries, it raises concern in the bias of data collected. These differences could be the result of systemic economic inequalities or cultural values that influence how happiness is both experienced and reported. To figure out the roots of this disparity, we created multiple scatter plots to determine whether GDP is a consistent predictor of happiness across all regions. By doing so, we hope to see whether there's evidence to suggest that happiness is valued different across regions.



For this graph, the strength of connection between GDP and happiness varies across different regions, that is, the strength of the link to happiness is not always positively correlated with the level of wealth in a region. For instance, we see a negative correlation with North America & ANZ despite there being positive associations in all other regions. This implies the importance of non-economic factors and that happiness can not be predicted by GDP alone.

With our data sorted into 10 regions, we created a violin plot to show how the distribution of all 8 predictors vary by region.

Key Predictors Distribution by Region



Within each subset, we see that the distribution of these key predictors vary considerably by region. For Log GDP per Capita, we see that Western Europe and North America & ANZ show higher averages than all other world regions. For Social Support, we see South Asia and Sub-Saharan Africa score considerably higher, on average, than all other world regions. For Freedom, we see South Asia, Southeast Asia, and Western Europe share relatively similar distributions among higher averages than the rest of the regions. High-income regions tend to have higher score across key predictors like GDP, social support, life expectancy, and freedom. Other regions show more variation and lower averages. This implies that disparities in these predictors contribute to global differences in happiness.

With so many predictors used to calculate Life Ladder, we suspect there to be interactions among these predictors that could be confounding. Thus, we created a correlation matrix to quantify how much the values relate to one another.

Table 3: Correlation Matrix of Happiness Indicators

	Life.Ladder	Log.GDP.per.capita	Social.support	Healthy.life.expectancy.at.birth	Freedom.to.make.life.choices	Generosity	Perceptions.of.corruption	Positive.affect	Negative.affect
Life.Ladder	1.000	0.788	0.723	0.724	0.530	0.170	-0.451	0.511	-0.338
Log.GDP.per.capita	0.788	1.000	0.696	0.831	0.355	-0.023	-0.350	0.230	-0.259
Social.support	0.723	0.696	1.000	0.601	0.403	0.060	-0.225	0.436	-0.451
Healthy.life.expectancy.at.birth	0.724	0.831	0.601	1.000	0.366	0.008	-0.305	0.220	-0.138
Freedom.to.make.life.choices	0.530	0.355	0.403	0.366	1.000	0.319	-0.481	0.587	-0.272
Generosity	0.170	-0.023	0.060	0.008	0.319	1.000	-0.282	0.320	-0.082
Perceptions.of.corruption	-0.451	-0.350	-0.225	-0.305	-0.481	-0.282	1.000	-0.292	0.277
Positive.affect	0.511	0.230	0.436	0.220	0.587	0.320	-0.292	1.000	-0.328
Negative.affect	-0.338	-0.259	-0.451	-0.138	-0.272	-0.082	0.277	-0.328	1.000

According to the tavel, economic, social, and health-related factors are the strongest contributors to happiness, while emotional and institutional trust variables play smaller roles.

Conclusion

While some factors, such as GDP per capita, social support, and life expectancy, are reliably linked to happiness worldwide, their proportional importance varies greatly by location, according to our analysis of the World Happiness Report 2023. While South Asia and Sub-Saharan Africa report lower averages, indicating deeper structural or cultural differences, Western Europe and North America & ANZ typically report higher Life Ladder scores, indicating generally greater life satisfaction.

Using exploratory visualizations, we found that while wealthier areas tend to have happier people, GDP is not a sufficient explanation for all of the differences in wellbeing. Particularly in areas with less economic development, social support and the perception of having the freedom to choose were found to be equally significant. These patterns were supported by our correlation study, which showed that happiness was strongly correlated with a number of variables other than financial standing.

These findings inform us that it is important to consider regional and cultural differences when talking about global welfare. While GDP and social support are important overall, their effects on happiness vary by region. In one part of the world, the elements that affect happiness may not have the same impact in another. As a result, a broad strategy is unlikely to succeed. It is more reasonable to use region-specific tactics that prioritize the most relevant factors for each site in order to improve life satisfaction across a variety of demographics.

Future research could improve this strategy even further by using multivariate modeling with regional interaction factors and possibly time series forecasting to assess the long-term effects of global events or policy changes on happiness trends.

Appendices

```
# break don of total values
Happiness_Data %>%
  summarize(
    total_rows = n(),
    total_years = n_distinct(year),
    total_countries = n_distinct(Country.name)
  ) %>% knitr::kable()
```

```
# Log-GDP
Happiness_Data %>%
  filter(is.finite(Log.GDP.per.capita)) %>%
  ggplot(data = Happiness_Data,
    mapping = aes(x = Log.GDP.per.capita)) +
  geom_histogram(binwidth = 0.2, na.rm = T,
    fill = 'lightblue4', color = 'black') +
  labs(
    title = "Distribution of Log GDP per Capita by Countries",
    x = "Log GDP per Capita",
    y = "Number of Countries"
  ) +
  theme_minimal()
```

```
# Social Support
Happiness_Data %>%
  filter(!is.na(Social.support)) %>%
  ggplot(data = Happiness_Data, mapping = aes(x = Social.support)) +
  geom_histogram(binwidth = 0.05, na.rm = T,
    fill = 'lightblue4', color = 'black') +
  labs(
    title = "Distribution of Social Support by Countries",
    x = "Social Support",
    y = "Number of Countries"
  ) +
  theme_minimal()
```

```
#Life Expectancy
Happiness_Data %>%
  filter(!is.na(Healthy.life.expectancy.at.birth)) %>%
  ggplot(data = Happiness_Data,
```

```

    mapping = aes(x = Healthy.life.expectancy.at.birth)) +
geom_histogram(binwidth = 4, na.rm = T,
               fill = 'lightblue4', color = 'black') +
labs(
  title = "Healthy Life Expectancy at Birth by Countries",
  x = "Healthy Life Expectancy at Birth",
  y = "Number of Countries"
) +
theme_minimal()

```

```

#Freedom of Choice
Happiness_Data %>%
  filter(!is.na(Freedom.to.make.life.choices)) %>%
  ggplot(data = Happiness_Data,
        mapping = aes(x = Freedom.to.make.life.choices)) +
  geom_histogram(binwidth = 0.05, na.rm = T,
                fill = 'lightblue4', color = 'black') +
  labs(
    title = "Freedom to make Life Choices by Countries",
    x = "Freedom Score",
    y = "Number of Countries"
  ) +
  theme_minimal()

```

```

# Generosity
Happiness_Data %>%
  filter(!is.na(Generosity)) %>%
  ggplot(data = Happiness_Data,
        mapping = aes(x = Generosity)) +
  geom_histogram(binwidth = 0.05, na.rm = T,
                fill = 'lightblue4', color = 'black') +
  labs(
    title = "Distribution of Generosity by Countries",
    x = "Generosity Score",
    y = "Number of Countries"
  ) +
  theme_minimal()

```

```

# Perceptions of corruption
Happiness_Data %>%
  filter(!is.na(Perceptions.of.corruption)) %>%
  ggplot(data = Happiness_Data,

```

```

    mapping = aes(x = Perceptions.of.corruption)) +
  geom_histogram(binwidth = 0.05, na.rm = T,
    fill = 'lightblue4', color = 'black') +
  labs(
    title = "Perceptions of Corruption by Countries",
    x = "Perceptions of Corruption Score",
    y = "Number of Countries"
  ) +
  theme_minimal()

```

```

# Positive Affect
Happiness_Data %>%
  filter(!is.na(Positive.affect)) %>%
  ggplot(data = Happiness_Data,
    mapping = aes(x = Positive.affect)) +
  geom_histogram(binwidth = 0.05, na.rm = T,
    fill = 'lightblue4', color = 'black') +
  labs(
    title = "Distribution of Positive Affect by Countries",
    x = "Positive Affect",
    y = "Number of Countries"
  ) +
  theme_minimal()

```

```

# Negative Affect
Happiness_Data %>%
  filter(!is.na(Negative.affect)) %>%
  ggplot(data = Happiness_Data,
    mapping = aes(x = Negative.affect)) +
  geom_histogram(binwidth = 0.05, na.rm = T,
    fill = 'lightblue4', color = 'black') +
  labs(
    title = "Distribution of Negative Affect by Countries",
    x = "Negative Affect",
    y = "Number of Countries"
  ) +
  theme_minimal()

```

```

# Ladder over time
Happiness_Data %>%
  group_by(year) %>%
  summarize(mean_life_ladder = mean(Life.Ladder, na.rm = TRUE)) %>%

```



```

ggplot(aes(x = year, y = mean_life_ladder)) +
  geom_line(color = "steelblue", linewidth = 1.2) +
  geom_point() +
  labs(title = "Global Average Life Ladder Score Over Time",
       x = "Year", y = "Average Life Ladder (Happiness)") +
  theme_minimal()

```

```

# World Region Seperation
Happiness_Data <- Happiness_Data %>%
  mutate(Region = case_when(

    Country.name %in% c("Austria", "Belgium", "Denmark", "Finland",
      "France", "Germany", "Iceland", "Ireland",
      "Italy", "Luxembourg", "Malta", "Netherlands",
      "Norway", "Portugal", "Spain", "Sweden",
      "Switzerland", "United Kingdom", "Greece", "Cyprus") ~
      "Western Europe",

    Country.name %in% c("Canada", "United States", "Australia", "New Zealand") ~
      "North America and ANZ",

    Country.name %in% c("Argentina", "Belize", "Bolivia", "Brazil",
      "Chile", "Colombia", "Costa Rica", "Cuba",
      "Dominican Republic", "Ecuador", "El Salvador", "Guatemala",
      "Haiti", "Honduras", "Jamaica", "Mexico",
      "Nicaragua", "Panama", "Paraguay", "Peru",
      "Suriname", "Trinidad and Tobago", "Uruguay", "Venezuela",
      "Guyana") ~ "Latin America and Caribbean",

    Country.name %in% c("Albania", "Bosnia and Herzegovina", "Bulgaria", "Croatia",
      "Czechia", "Estonia", "Hungary", "Kosovo",
      "Latvia", "Lithuania", "Montenegro", "North Macedonia",
      "Poland", "Romania", "Serbia", "Slovakia",
      "Slovenia") ~ "Central and Eastern Europe",

    Country.name %in% c("Armenia", "Azerbaijan", "Belarus", "Kazakhstan",
      "Kyrgyzstan", "Moldova", "Russia", "Tajikistan",
      "Turkmenistan", "Ukraine", "Uzbekistan", "Georgia")
    ~ "CIS",

    Country.name %in% c("Angola", "Benin", "Botswana", "Burkina Faso",
      "Burundi", "Cameroon", "Central African Republic", "Chad",

```

```

"Congo (Brazzaville)", "Congo (Kinshasa)", "Eswatini", "Ethiopia",
"Gabon", "Gambia", "Ghana", "Guinea",
"Ivory Coast", "Kenya", "Lesotho", "Liberia",
"Madagascar", "Malawi", "Mali", "Mauritania",
"Mauritius", "Mozambique", "Namibia", "Niger",
"Nigeria", "Rwanda", "Senegal", "Sierra Leone",
"Somalia", "South Africa", "South Sudan", "Sudan",
"Togo", "Uganda", "Tanzania", "Zambia",
"Zimbabwe", "Somaliland region", "Comoros", "Djibouti") ~
"Sub-Saharan Africa",

Country.name %in% c("Afghanistan", "Bangladesh", "Bhutan", "India",
"Nepal", "Pakistan", "Sri Lanka", "Maldives") ~
"South Asia",

Country.name %in% c("China", "Hong Kong S.A.R. of China", "Japan", "Mongolia",
"South Korea", "Taiwan Province of China") ~
"East Asia",

Country.name %in% c("Cambodia", "Indonesia", "Laos", "Malaysia",
"Myanmar", "Philippines", "Singapore", "Thailand",
"Vietnam") ~ "Southeast Asia",

Country.name %in% c("Algeria", "Bahrain", "Egypt", "Iran",
"Iraq", "Israel", "Jordan", "Kuwait",
"Lebanon", "Libya", "Morocco", "Oman",
"Qatar", "Saudi Arabia", "State of Palestine", "Syria",
"Tunisia", "Turkiye", "United Arab Emirates", "Yemen") ~
"Middle East and North Africa",

TRUE ~ "Other"
))

```

```

#Ladder Score Distribution
Happiness_Data %>%
  drop_na(Life.Ladder, Region) %>%
  ggplot(aes(x = Region, y = Life.Ladder, fill = Region)) +
  geom_boxplot() +
  labs(title = "Distribution of Life Ladder Scores by Region",
       x = "Region", y = "Life Ladder (Happiness)",
       fill = "Region") +
  scale_fill_brewer(palette = 'Set3') +

```

```
theme_minimal() +
theme(axis.text.x = element_blank(),
      legend.position = "right")
```

```
#Happines vs GDP by Region
Happiness_Data <- Happiness_Data %>%
  mutate(Region_label = case_when(
    Region == "Latin America and Caribbean" ~ "LatAm & Caribbean",
    Region == "North America and ANZ" ~ "NA & ANZ",
    Region == "Middle East and North Africa" ~ "MENA",
    Region == "Central and Eastern Europe" ~ "Central & East Europe",
    TRUE ~ Region
  ))

Happiness_Data %>%
  drop_na(Life.Ladder, Log.GDP.per.capita, Region_label) %>%
  ggplot(aes(x = Log.GDP.per.capita, y = Life.Ladder)) +
  geom_point(alpha = 0.4, size = 1.2, color = 'darkslategray') +
  geom_smooth(method = "lm", color = "tomato", se = FALSE, linewidth = 1) +
  facet_wrap(~ Region_label, ncol = 2, scales = 'free_y') +
  labs(
    title = "Happiness vs Log GDP per Capita by Region",
    x = "Log GDP per Capita",
    y = "Life Ladder"
  ) +
  theme_minimal() +
  theme(
    strip.text = element_text(size = 10, face = "bold"),
    plot.title = element_text(hjust = 0.5, size = 14, face = "bold")
  )
```

```
# Predictor distribution by Region
library(ggplot2)
library(dplyr)
library(tidyr)
Happiness_Data %>%
  drop_na(Log.GDP.per.capita, Social.support,
    Healthy.life.expectancy.at.birth,
    Freedom.to.make.life.choices, Generosity,
    Perceptions.of.corruption,
    Positive.affect, Negative.affect, Region) %>%
  pivot_longer(cols = c(
```

```

Log.GDP.per.capita, Social.support, Healthy.life.expectancy.at.birth,
Freedom.to.make.life.choices, Generosity, Perceptions.of.corruption,
Positive.affect, Negative.affect),
names_to = "Variable", values_to = "Value") %>%

# Shorten Region names & add line breaks
mutate(Region = recode(Region,
  "Latin America and Caribbean" = "Latin America\nand Caribbean",
  "North America and ANZ" = "NA\n& ANZ",
  "Middle East and North Africa" = "MENA",
  "Central and Eastern Europe" = "Central &\nEastern Europe",
  "Western Europe" = "Western\nEurope" # ←
)) %>%

# Shorten variable names
mutate(Variable = recode(Variable,
  "Log.GDP.per.capita" = "Log GDP",
  "Social.support" = "Social Support",
  "Healthy.life.expectancy.at.birth" = "Life Expectancy",
  "Freedom.to.make.life.choices" = "Freedom")
) %>%

ggplot(aes(x = Region, y = Value, fill = Region)) +
geom_violin(trim = FALSE, alpha = 0.7) +
facet_wrap(~ Variable, scales = "free", ncol = 3) +
theme_minimal() +
guides(fill = guide_legend(nrow = 2, byrow = TRUE)) +
theme(
  legend.position = "bottom",
  legend.box = "horizontal",
  legend.title = element_blank(),
  legend.text = element_text(size = 8),
  legend.key.size = unit(0.5, "lines"),
  legend.spacing.x = unit(0.3, 'cm'),
  legend.margin = margin(t = -5, unit = "pt"),
  axis.text.x = element_blank(),          # remove x-axis labels
  axis.ticks.x = element_blank(),         # remove x-axis ticks
  strip.text = element_text(size = 5.5, face = "bold"), # smaller facet titles
  plot.title = element_text(hjust = 0.5, size = 14, face = "bold"),
  plot.margin = margin(t = 10, r = 10, b = 10, l = 10) # prevent clipping
) +
labs(

```

```

    title = "Key Predictors Distribution by Region",
    x = "Region", y = "Values Scored"
  )

```

```

#Correlation Matrix

```

```

library(dplyr)

```

```

library(knitr)

```

```

library(kableExtra)

```

```

Happiness_Data %>%

```

```

  select(

```

```

    Life.Ladder,

```

```

    Log.GDP.per.capita,

```

```

    Social.support,

```

```

    Healthy.life.expectancy.at.birth,

```

```

    Freedom.to.make.life.choices,

```

```

    Generosity,

```

```

    Perceptions.of.corruption,

```

```

    Positive.affect,

```

```

    Negative.affect

```

```

  ) %>%

```

```

  drop_na() %>%

```

```

  cor() %>%

```

```

  round(3) %>%

```

```

  kable("latex", booktabs = TRUE, caption = "Correlation Matrix of Happiness Indicators")%>%

```

```

  kable_styling(latex_options = c("scale_down", "hold_position"))

```