

# Trends and Factors Influencing Global Happiness - World Happiness Record 2023

Jeremy Djohar Riyadi - 2702219572

Joseph Adinata Elkos - 2702224282

Justin Hartanto Widjaja - 2702222610

Liliana Djaja Witama - 2702219774

Group 6:

2024-06-22

## Introduction

Happiness is a crucial indicator of a country's quality of life. Furthermore, with the raise of mental health issues around the world, we believe it is important to improve our life condition so that it doesn't affect our quality of life. In the context of public policy, economics, and public health, understanding the factors that influence happiness can help governments and organizations design more effective policies, which is essential to achieve one of the SDGs, mainly SDG 3: Good Health and Well-being.

This report aims to analyze the World Happiness Report 2023, one of the most comprehensive and well-known happiness surveys globally. This report (WHR2023) presents data from various countries with rigorous methods and reliable data sources. Moreover, it covers various relevant variables such as GDP per capita, social support, healthy life expectancy, freedom to make life choices, generosity, and perceptions of corruption. By using this dataset, our analysis can provide deep and broad insights into the factors influencing global happiness.

The main topic of this report is the analysis of trends and factors influencing global happiness. We will explore how economic, social, and health variables contribute to happiness and how regional differences affect these outcomes. The primary goals of this analysis are to answer several key questions:

- What are the main factors influencing happiness in various countries?
- What is the relationship between GDP per capita and happiness?
- Are there significant differences in happiness levels based on regions?
- How do social support and healthy life expectancy contribute to happiness?
- Where does Indonesia stand compared to other countries?

In this report, we will use Exploratory Data Analysis (EDA) methods to understand the distribution and relationships between variables. Statistical analysis and data visualization will also be used to illustrate the main findings.

This analysis is highly relevant in various fields. From a public policy perspective, the results can be used to design better programs to enhance societal well-being. Economically, understanding the relationship between GDP and happiness can aid in decision-making regarding investment and resource distribution. In public health, this analysis can provide insights into the importance of physical and mental health in enhancing happiness.

This report is intended for researchers, policymakers, and other stakeholders interested in the factors influencing societal happiness. Additionally, this report is relevant to international organizations, academics, and the general public who wish to gain a deeper understanding of the factors affecting global happiness.

This analysis is based on data available in the World Happiness Report 2023, which is the result of surveys and measurements from various sources. Some assumptions made are that this data is accurate and representative of the broader population. Limitations of the analysis include potential biases in the survey and the challenge of objectively measuring happiness.

This analysis is meaningful because happiness is one of the main goals in human life. By understanding the factors that influence happiness, we hope to provide valuable insights for all parties so that we can work towards a happier, more equitable, more productive, and more prosperous world.

## Data Description

Dataset used: World Happiness Report 2023, a partnership of Gallup, the Oxford Wellbeing Research Centre, the UN Sustainable Development Solutions Network, and the WHR's Editorial Board.

Source: Helliwell, J. F., Layard, R., Sachs, J. D., Aknin, L. B., De Neve, J.-E., & Wang, S. (Eds.). (2023). World Happiness Report 2023 (11th ed.). Sustainable Development Solutions Network.

World Happiness Report 2023, one of the most comprehensive and well-known happiness surveys globally. This report presents data from various countries with rigorous methods and reliable data sources. Moreover, it covers various relevant variables such as GDP per capita, social support, healthy life expectancy, freedom to make life choices, generosity, and perceptions of corruption, which are related to countries' happiness score. Description of mentioned variables are as follows:

- **country:** name of the country
- **region:** region or continent geographically
- **happiness\_score:** overall happiness measure
- **gdp\_per\_capita:** measure of gross domestic product per capita

- **social\_support**: measure of the quality of social support provided by the country and an individual's social life
- **healthy\_life\_expectancy**: measure of an individual's healthy life expectancy
- **freedom\_to\_make\_life\_choices**: individual's freedom to make decisions
- **generosity**: measure of an individual's generosity
- **perceptions\_of\_corruption**: individual's perception of corruption in a country

## Data Preprocessing

### Calling libraries and functions used

```
source("explorationFunction.R")
library(corrplot)
library(ggplot2)
library(dplyr)
library(plotly)
```

### Opening csv file

```
x = read.csv("C:/Users/lenovo/OneDrive - Bina Nusantara/SEMESTER 2/Data
Mining and Visualization/assignment 1/WHR_2023.csv")
df = data.frame(x)
```

Opens a csv file, then converts it to a data frame, and store it in a variable called df.

### Viewing variable characters

```
str(df)

## 'data.frame':    137 obs. of  9 variables:
## $ country          : chr  "Finland" "Denmark" "Iceland"
## $ region           : chr  "Western Europe" "Western Europe"
## $ happiness_score   : num  7.8 7.59 7.53 7.47 7.4 ...
## $ gdp_per_capita    : num  1.89 1.95 1.93 1.83 1.94 ...
## $ social_support    : num  1.58 1.55 1.62 1.52 1.49 ...
## $ healthy_life_expectancy : num  0.535 0.537 0.559 0.577 0.545 0.562
## $ freedom_to_make_life_choices: num  0.772 0.734 0.738 0.569 0.672 0.754
## $ generosity        : num  0.126 0.208 0.25 0.124 0.251 0.225
## $ perceptions_of_corruption : num  0.535 0.525 0.187 0.158 0.394 0.52
```

The data frame has 137 observations with 9 variables, 2 categorical and 7 numerical. The two categorical variables, namely country and region, have character data types. Meanwhile, the numerical variables consist of happiness\_score, gdp\_per\_capita, social\_support, healthy\_life\_expectancy, freedom\_to\_make\_life\_choices, generosity, and perceptions\_of\_corruption. The first 5 observations for each variable in the data frame are also displayed.

```
summary(df)
```

```
##      country          region      happiness_score gdp_per_capita
## Length:137      Length:137      Min.   :1.859      Min.   :0.000
## Class :character Class :character 1st Qu.:4.724      1st Qu.:1.099
## Mode  :character Mode  :character Median :5.684      Median :1.449
##                                     Mean  :5.540      Mean  :1.407
##                                     3rd Qu.:6.334      3rd Qu.:1.798
##                                     Max.   :7.804      Max.   :2.200
##
## social_support healthy_life_expectancy freedom_to_make_life_choices
## Min.   :0.000      Min.   :0.0000      Min.   :0.000
## 1st Qu.:0.962      1st Qu.:0.2485      1st Qu.:0.455
## Median :1.227      Median :0.3895      Median :0.557
## Mean   :1.156      Mean   :0.3662      Mean   :0.540
## 3rd Qu.:1.401      3rd Qu.:0.4875      3rd Qu.:0.656
## Max.   :1.620      Max.   :0.7020      Max.   :0.772
##                                     NA's   :1
## generosity      perceptions_of_corruption
## Min.   :0.0000      Min.   :0.0000
## 1st Qu.:0.0970      1st Qu.:0.0600
## Median :0.1370      Median :0.1110
## Mean   :0.1485      Mean   :0.1459
## 3rd Qu.:0.1990      3rd Qu.:0.1870
## Max.   :0.4220      Max.   :0.5610
##
```

Displays summary data in the dataframe, in the form of the mean, 5 number summary, and the number of missing values(NA). It can be seen that for the variable healthy\_life\_expectancy there is 1 null value.

### Checking and deleting null values

```
sum(is.na(df))
```

```
## [1] 1
```

```
which(!complete.cases(df))
```

```
## [1] 99
```

```
df <- df[-99, ]
```

```
rownames(df) <- seq_len(nrow(df))
```

We checked for NA values in the dataset. NA values or missing values occur when no data value is stored for the variable in observation. It is important to ensure that the dataset used is complete so that the analysis will be reliable. In this dataset, there is 1 null value in row 99 column 5, which is then deleted because its number is insignificant. The data frame now has 136 rows and 9 columns.

### Checking for duplicated values

```
sum(duplicated(df))
```

```
## [1] 0
```

We checked if there is any duplicated values in the dataset. A duplicate value is one in which all values in at least one row are identical to all of the values in another row. It is important to check for duplicated values to avoid inconsistencies and maintain accuracy within a dataset. There is no duplicated values in this dataframe.

### Looking for outliers in each variable

An outlier is a data point that differs significantly from other observations. It is important to look for outliers in a dataset because it may affect the validity and reliability of statistical tests that may be performed.

```
FindOutliers(df$happiness_score)
```

```
## $summary
##      method    n nMiss nOut   lowLim   upLim minNom maxNom
## 1 ThreeSigma 136     0    1 2.115920 8.972963  2.392  7.804
## 2      Hampel 136     0    1 2.288709 9.098291  2.392  7.804
## 3 BoxplotRule 136     0    9 3.882500 8.802500  3.982  7.804
##
## $threeSigma
##   index values type
## 1   136  1.859    L
##
## $Hampel
##   index values type
## 1   136  1.859    L
##
## $boxplotRule
##   index values type
## 1   128  3.694    L
## 2   129  3.545    L
## 3   130  3.495    L
## 4   131  3.435    L
## 5   132  3.207    L
## 6   133  3.204    L
## 7   134  3.138    L
## 8   135  2.392    L
## 9   136  1.859    L
```

According to the Three Sigma method there is 1 outlier. According to the Hampel method, there is 1 outlier. According to the Boxplot Rule there are 9 outliers.

```
FindOutliers(df$gdp_per_capita)
```

```
## $summary
##      method    n nMiss nOut   lowLim   upLim minNom maxNom
## 1 ThreeSigma 136     0    1  0.1070118 2.710826  0.531  2.2
## 2      Hampel 136     0    0 -0.1119017 3.014902  0.000  2.2
## 3 BoxplotRule 136     0   10  0.7476250 2.848375  0.758  2.2
```

```
##
## $threeSigma
##   index values type
## 1    88      0     L
##
## $Hampel
## [1] index values type
## <0 rows> (or 0-length row.names)
##
## $boxplotRule
##   index values type
## 1    88 0.000     L
## 2    97 0.570     L
## 3   108 0.561     L
## 4   113 0.622     L
## 5   124 0.628     L
## 6   126 0.632     L
## 7   130 0.637     L
## 8   132 0.531     L
## 9   134 0.670     L
## 10  136 0.645     L
```

According to the Three Sigma method there is 1 outlier. According to the Hampel method there are 0 outliers. According to the Boxplot Rule there are 10 outliers.

```
FindOutliers(df$social_support)
```

```
## $summary
##      method   n nMiss nOut   lowLim   upLim minNom maxNom
## 1 ThreeSigma 136     0    1 0.1732997 2.136877 0.242 1.62
## 2      Hampel 136     0    2 0.2892381 2.161762 0.327 1.62
## 3 BoxplotRule 136     0   18 0.7390000 2.063500 0.756 1.62
##
## $threeSigma
##   index values type
## 1   136      0     L
##
## $Hampel
##   index values type
## 1   115 0.242     L
## 2   136 0.000     L
##
## $boxplotRule
##   index values type
## 1    86 0.665     L
## 2    93 0.584     L
## 3    99 0.535     L
## 4   101 0.727     L
## 5   107 0.657     L
## 6   108 0.628     L
```

```
## 7      115  0.242    L
## 8      117  0.513    L
## 9      118  0.614    L
## 10     119  0.637    L
## 11     121  0.642    L
## 12     124  0.644    L
## 13     125  0.674    L
## 14     129  0.327    L
## 15     130  0.479    L
## 16     134  0.540    L
## 17     135  0.476    L
## 18     136  0.000    L
```

According to the Three Sigma method there is 1 outlier. According to the Hampel method, there are 2 outliers. According to the Boxplot Rule there are 18 outliers.

```
FindOutliers(df$healthy_life_expectancy)
```

```
## $summary
##      method    n nMiss nOut      lowLim      upLim minNom maxNom
## 1 ThreeSigma 136     0     0 -0.1038968 0.8362498  0.000  0.702
## 2      Hampel 136     0     0 -0.1308926 0.9098926  0.000  0.702
## 3 BoxplotRule 136     0    15  0.1290000 0.8460000  0.137  0.702
##
## $threeSigma
## [1] index values type
## <0 rows> (or 0-length row.names)
##
## $Hampel
## [1] index values type
## <0 rows> (or 0-length row.names)
##
## $boxplotRule
##      index values type
## 1      91  0.072    L
## 2      93  0.120    L
## 3      95  0.092    L
## 4      96  0.118    L
## 5      97  0.000    L
## 6     103  0.107    L
## 7     113  0.043    L
## 8     115  0.124    L
## 9     119  0.106    L
## 10     127  0.095    L
## 11     131  0.087    L
## 12     132  0.105    L
## 13     133  0.069    L
## 14     134  0.092    L
## 15     136  0.087    L
```

According to the Three Sigma method there are 0 outliers. According to the Hampel method there are 0 outliers. According to the Boxplot Rule there are 15 outliers.

```
FindOutliers(df$freedom_to_make_life_choices)
```

```
## $summary
##      method    n nMiss nOut      lowLim      upLim minNom maxNom
## 1 ThreeSigma 136     0    1 0.09189811 0.9899254  0.117  0.772
## 2      Hampel 136     0    1 0.11272000 1.0022800  0.117  0.772
## 3 BoxplotRule 136     0   13 0.35975000 0.9537500  0.363  0.772
##
## $threeSigma
##   index values type
## 1   136      0    L
##
## $Hampel
##   index values type
## 1   136      0    L
##
## $boxplotRule
##   index values type
## 1     58 0.248    L
## 2     81 0.252    L
## 3     94 0.346    L
## 4     98 0.351    L
## 5    100 0.281    L
## 6    102 0.320    L
## 7    105 0.125    L
## 8    109 0.259    L
## 9    118 0.286    L
## 10   126 0.187    L
## 11   129 0.117    L
## 12   135 0.123    L
## 13   136 0.000    L
```

According to the Three Sigma method there is 1 outlier. According to the Hampel method, there is 1 outlier. According to the Boxplot Rule there are 13 outliers.

```
FindOutliers(df$generosity)
```

```
## $summary
##      method    n nMiss nOut      lowLim      upLim minNom maxNom
## 1 ThreeSigma 136     0    2 -0.07889173 0.3770682  0.000  0.332
## 2      Hampel 136     0    2 -0.08711390 0.3621139  0.000  0.332
## 3 BoxplotRule 136     0   12  0.04812500 0.3503750  0.053  0.332
##
## $threeSigma
##   index values type
## 1    84 0.422    U
## 2   116 0.400    U
##
```



```
## $Hampel
##   index values type
## 1    84  0.422    U
## 2   116  0.400    U
##
## $boxplotRule
##   index values type
## 1    24  0.044    L
## 2    47  0.009    L
## 3    56  0.031    L
## 4    58  0.008    L
## 5    90  0.000    L
## 6    94  0.036    L
## 7    99  0.013    L
## 8   109  0.022    L
## 9   120  0.038    L
## 10   131  0.021    L
## 11    84  0.422    U
## 12   116  0.400    U
```

According to the Three Sigma method there are 2 outliers. According to the Hampel method, there are 2 outliers. According to the Boxplot Rule there are 12 outliers.

```
FindOutliers(df$perceptions_of_corruption)
```

```
## $summary
##      method   n nMiss nOut   lowLim   upLim minNom maxNom
## 1 ThreeSigma 136     0    2 -0.2345482 0.5275041     0  0.525
## 2 Hampel     136     0   12 -0.1437485 0.3677485     0  0.365
## 3 BoxplotRule 136     0   11 -0.0045000 0.3810000     0  0.372
##
## $threeSigma
##   index values type
## 1     1  0.535    U
## 2    25  0.561    U
##
## $Hampel
##   index values type
## 1     1  0.535    U
## 2     2  0.525    U
## 3     5  0.394    U
## 4     6  0.520    U
## 5     7  0.463    U
## 6     8  0.475    U
## 7     9  0.418    U
## 8    10  0.471    U
## 9    14  0.409    U
## 10   25  0.561    U
## 11   31  0.372    U
## 12   82  0.390    U
```

```
##
## $boxplotRule
##      index values type
## 1         1  0.535    U
## 2         2  0.525    U
## 3         5  0.394    U
## 4         6  0.520    U
## 5         7  0.463    U
## 6         8  0.475    U
## 7         9  0.418    U
## 8        10  0.471    U
## 9        14  0.409    U
## 10       25  0.561    U
## 11       82  0.390    U
```

According to the Three Sigma method there are 2 outliers. According to the Hampel method there are 12 outliers. According to the Boxplot Rule there are 11 outliers.

### Removing outliers?

In deciding whether the outliers should be removed or not, we considered a couple points. First of all, this dataset has a small number of observations, of only 136 observation. Hence, each observation represents the dataset's natural variability. That's why from this point of view, we shouldn't remove the outliers so that we won't reduce the dataset's reliability.

```
#Z score method
sd(df$happiness_score)

## [1] 1.142841

threshold <- sd(df$happiness_score) * 3

dfCleaned <- df %>%
  filter(abs(happiness_score - mean(happiness_score)) < threshold)

sd(dfCleaned$happiness_score)

## [1] 1.101689
```

From statistical point of view, we also did a test on how far the standard deviation of the happiness score variable change after the outliers are removed. Before the outliers are removed, the standard deviation is 1.142841. After the outliers are removed, the standard deviation becomes 1.101689. Because the difference is not significant, the outliers are not needed to be removed.

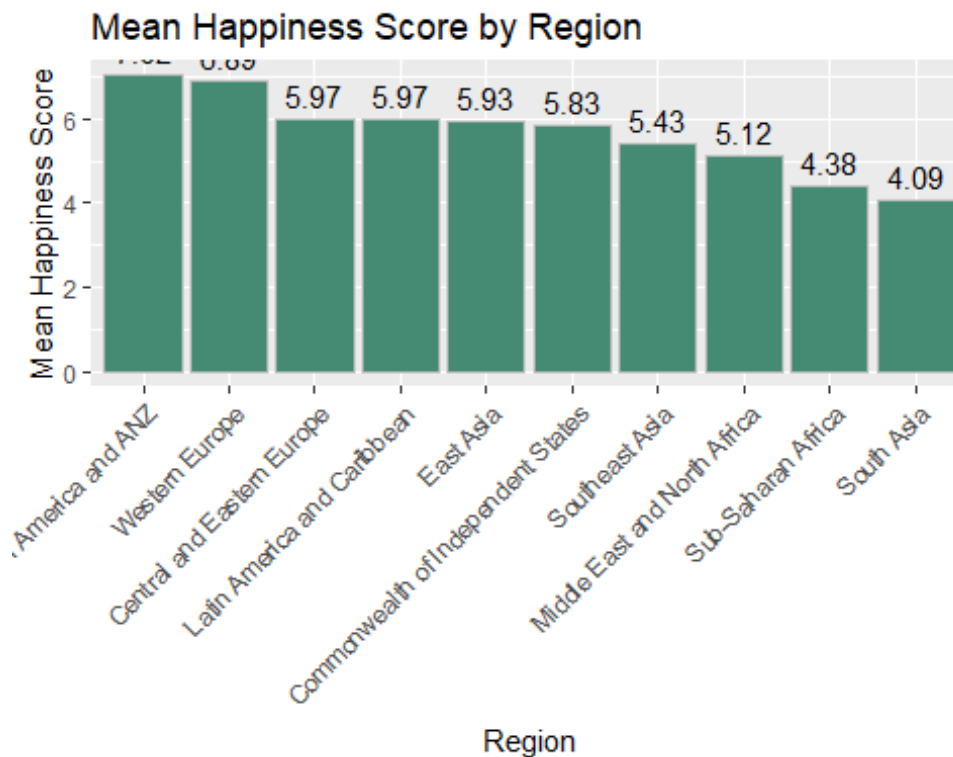
### Data Exploration

We performed data exploration using the Exploratory Data Analysis (EDA) method. EDA is an approach of analyzing data sets to summarize their main characteristics, often using

statistical graphics and other data visualization methods. We used this method to help us gain insights, identify patterns, and communicate findings easier.

#### Average happiness score by region

```
happinessByRegion <- df %>%  
  group_by(region) %>%  
  summarize(mean_happiness = mean(happiness_score, na.rm = TRUE))  
  
ggplot(happinessByRegion,  
  aes(x = reorder(region, -mean_happiness), y = mean_happiness)) +  
  geom_bar(stat = "identity", fill = "#458B74", color = "gray") +  
  geom_text(aes(label = round(mean_happiness, 2)), vjust = -0.5) +  
  labs(x = "Region", y = "Mean Happiness Score", title = "Mean Happiness  
Score by Region") +  
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

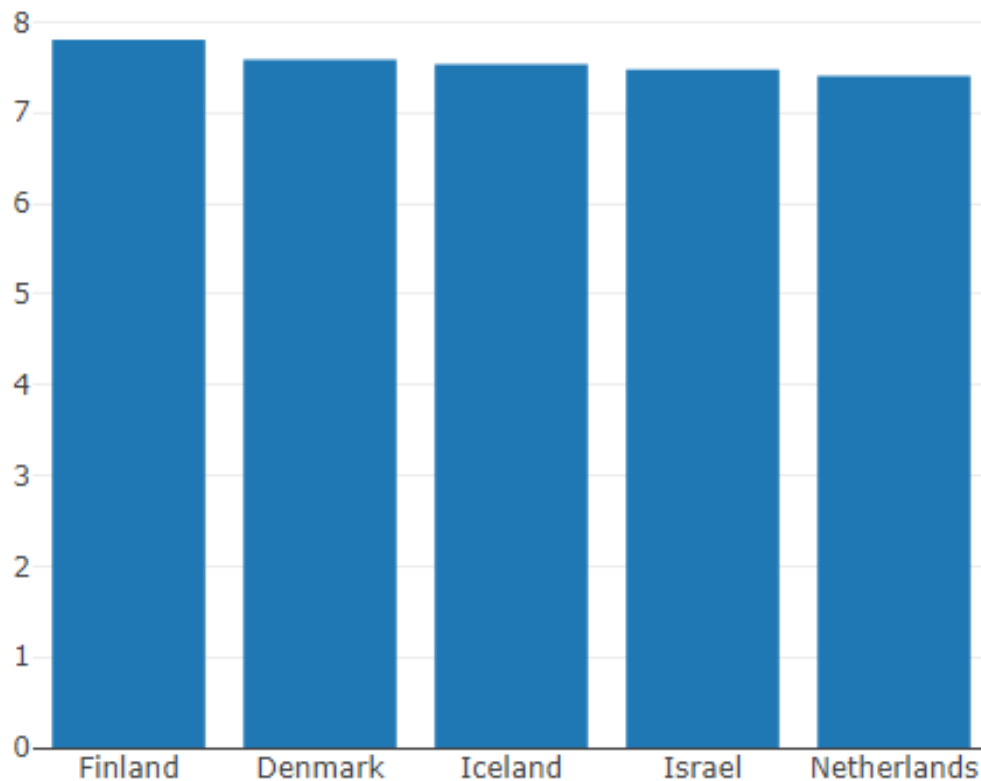


The bar plot shows a comparison between the average happiness\_score per region. It was found that North America and ANZ (Australia and New Zealand) had the highest happiness\_score of 7.02.

#### Top 5 happiest countries

```
data_sorted <- df[order(-df$happiness_score), ]  
  
top_5_happy <- data_sorted[1:5, ]
```

```
plot_ly(x = factor(top_5_happy$country, levels = top_5_happy$country),  
        y = top_5_happy$happiness_score,  
        type = 'bar',  
)
```



Here is a simple bar graph on the top 5 countries with the highest happiness level in 2023. We can see that Finland is the happiest country in 2023. We can see its happiness level when we hover through its bar, which is 7.804. Followed by Denmark, with a happiness level of 7.586, Iceland with a happiness level of 7.53, Israel with a happiness level of 7.473, and rounding the top 5, Netherlands with a happiness level of 7.403.

### Statistical Analysis

Statistical analysis is the science of collecting, exploring and presenting large amounts of data to discover underlying patterns and trends. We performed a correlation test among numerical variables, which is the happiness score variable and factors that may influence happiness score. We did this test because we wanted to know which factor variables correlate with happiness score the most, how strongly they are correlated, and the direction of the correlation.

### Looking at correlations between numerical variables

```
newDf <- subset(df, select = -c(region, country))
```

```
c <- cor(newDf)
```

```
corrplot(c, type="upper", method="number")
```



This correlation plot displays the correlation between variables. The darker the color and the closer the number is to -1.0 or 1.0, the greater the correlation between the variables. For example, the correlation between happiness\_score and gdp\_per\_capita is 0.78. Because 0.78 is close to 1.0, happiness\_score and gdp\_per\_capita are strongly correlated. Meanwhile, the correlation between healthy\_life\_expectancy and generosity is only 0.14, therefore the correlation is weak. It can be concluded that the happiness\_score variable has a strong correlation with gdp\_per\_capita, social\_support, healthy\_life\_expectancy, and freedom\_to\_make\_life\_choices.

### Creating a scatter plot to show the relationship between variables

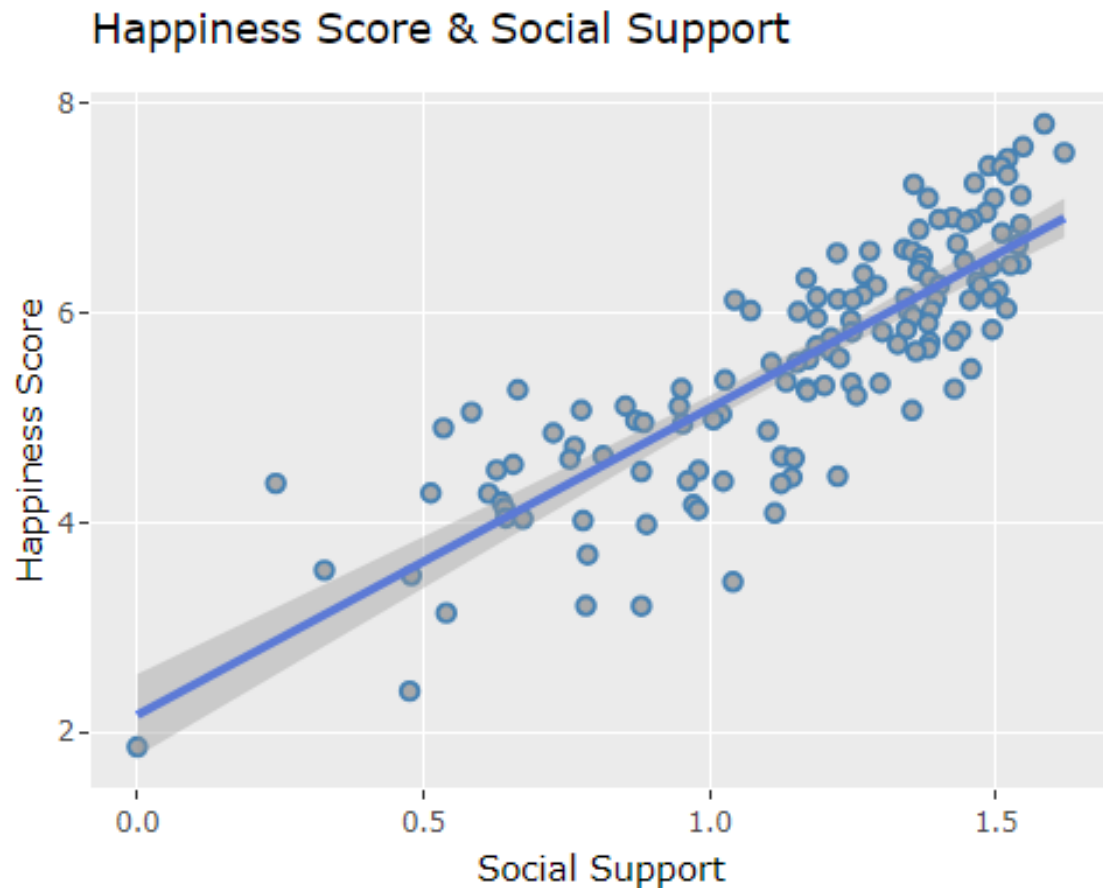
```
hgdp <- ggplot(data=df, mapping=aes(x=gdp_per_capita,y=happiness_score)) +  
  geom_point(pch=21,color="steelblue",fill="darkgray",size=2) +  
  labs(x="GDP",y="Happiness Score", title = "Happiness Score & GDP") +  
  geom_smooth(method="lm")
```

```
ggplotly(hgdp)
```



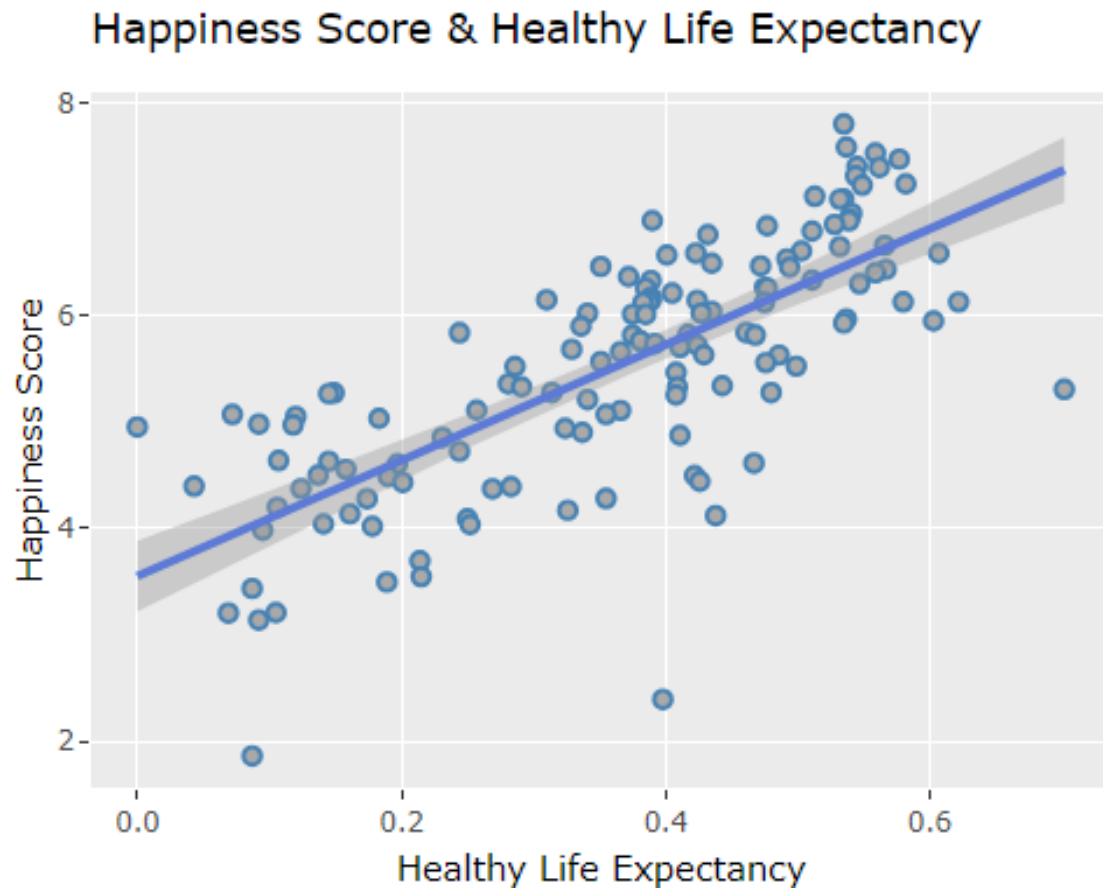
happiness\_score and gdp\_per\_capita have a linear relationship. The higher the happiness\_score, the higher the gdp\_per\_capita.

```
hs <- ggplot(data=df, mapping=aes(x=social_support,y=happiness_score)) +  
  geom_point(pch=21,color="steelblue",fill="darkgray",size=2) +  
  labs(x="Social Support",y="Happiness Score", title = "Happiness Score &  
Social Support") +  
  geom_smooth(method="lm")  
  
ggplotly(hs)
```



happiness\_score and social\_support have a linear relationship. The higher the happiness\_score, the higher the social\_support.

```
hh <- ggplot(data=df,  
mapping=aes(x=healthy_life_expectancy,y=happiness_score)) +  
  geom_point(pch=21,color="steelblue",fill="darkgray",size=2) +  
  labs(x="Healthy Life Expectancy",y="Happiness Score", title = "Happiness  
Score & Healthy Life Expectancy") +  
  geom_smooth(method="lm")  
  
ggplotly(hh)
```

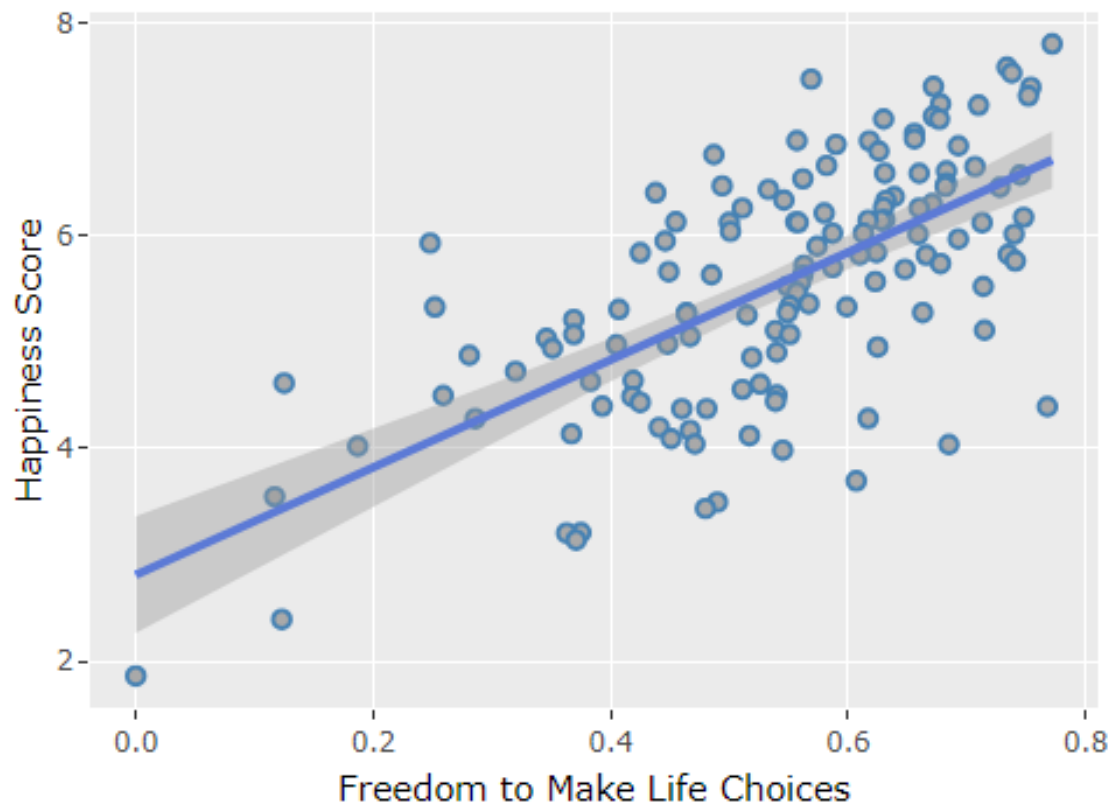


happiness\_score and healthy\_life\_expectancy have a linear relationship. The higher the happiness\_score, the higher the healthy\_life\_expectancy.

```
hf <- ggplot(data=df,  
mapping=aes(x=freedom_to_make_life_choices,y=happiness_score)) +  
  geom_point(pch=21,color="steelblue",fill="darkgray",size=2) +  
  labs(x="Freedom to Make Life Choices",y="Happiness Score", title =  
"Happiness Score & Freedom to Make Life Choices") +  
  geom_smooth(method="lm")  
  
ggplotly(hf)
```

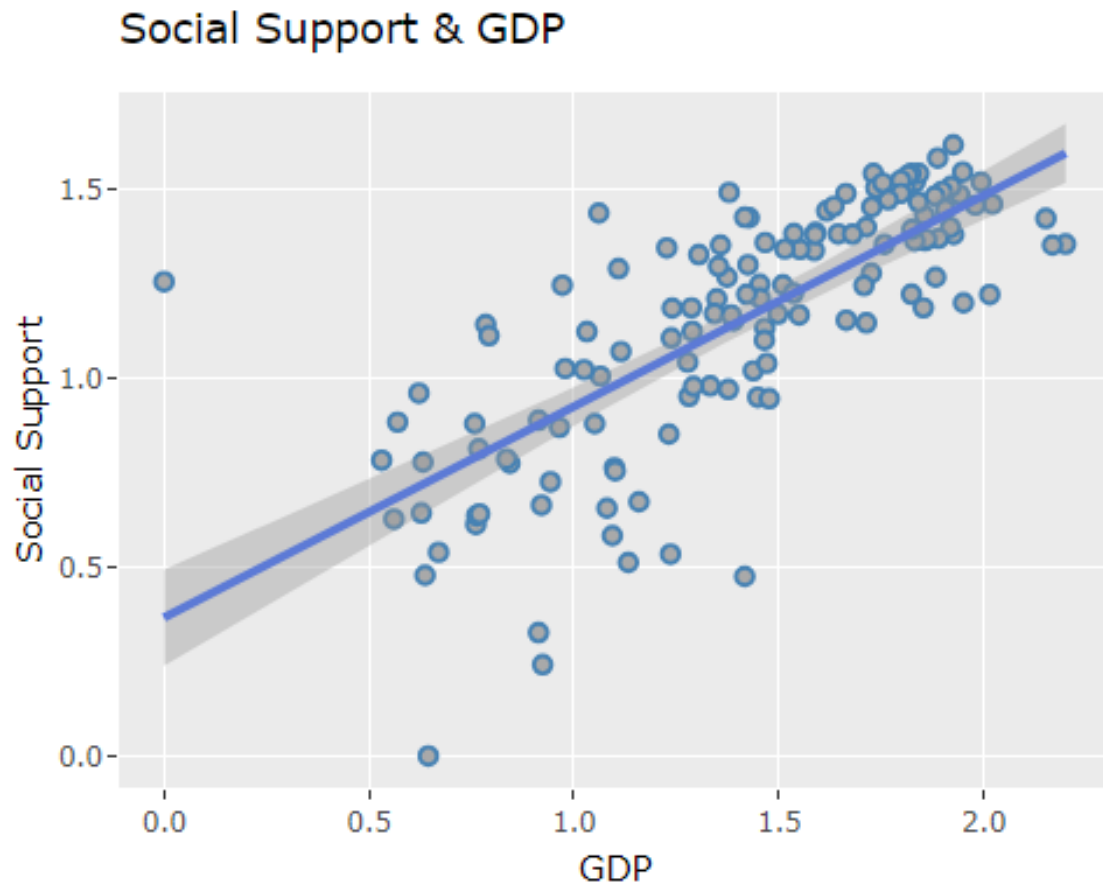


## Happiness Score & Freedom to Make Life Choices



happiness\_score and freedom\_to\_make\_life\_choices have a linear relationship. The higher the happiness\_score, the higher the freedom\_to\_make\_life\_choices.

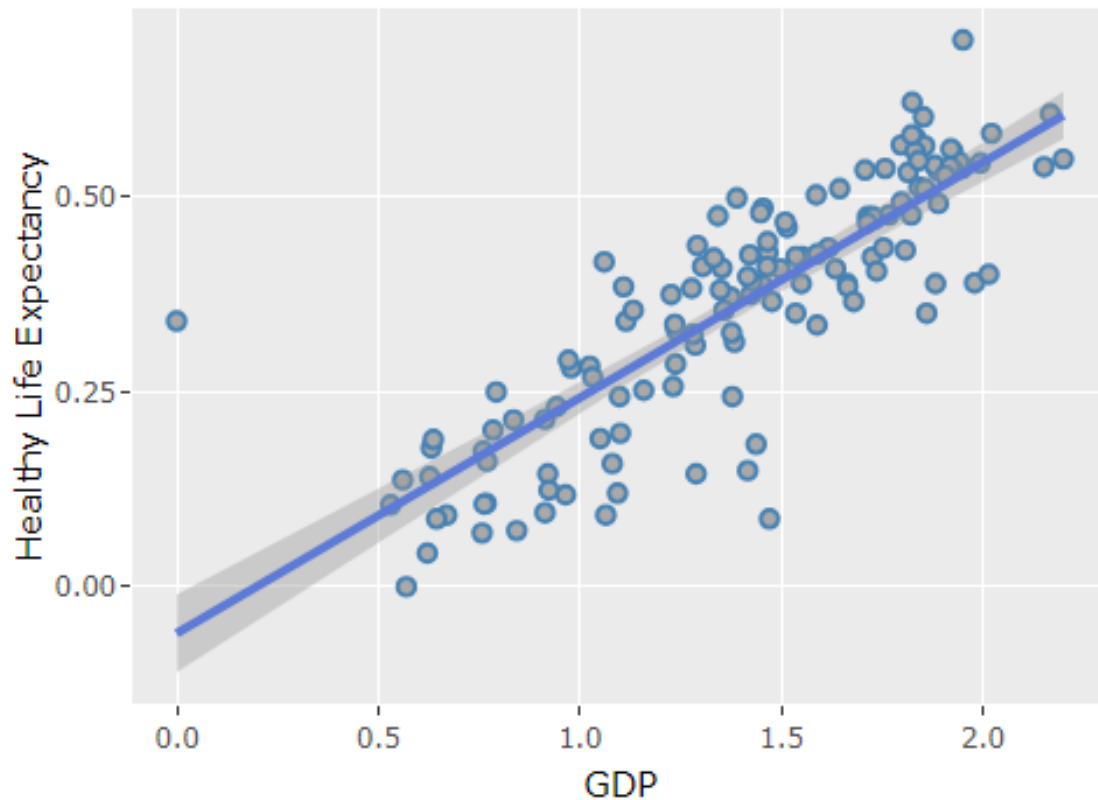
```
sg <- ggplot(data=df, mapping=aes(x=gdp_per_capita,y=social_support)) +  
  geom_point(pch=21,color="steelblue",fill="darkgray",size=2) +  
  labs(x="GDP",y="Social Support", title = "Social Support & GDP") +  
  geom_smooth(method="lm")  
  
ggplotly(sg)
```



gdp\_per\_capita and social\_support have a linear relationship. The higher gdp\_per\_capita, the higher social\_support.

```
hg <- ggplot(data=df,  
mapping=aes(x=gdp_per_capita,y=healthy_life_expectancy)) +  
  geom_point(pch=21,color="steelblue",fill="darkgray",size=2) +  
  labs(x="GDP",y="Healthy Life Expectancy", title = "Healthy Life Expectancy  
& GDP") +  
  geom_smooth(method="lm")  
  
ggplotly(hg)
```

## Healthy Life Expectancy & GDP



`gdp_per_capita` and `healthy_life_expectancy` have a linear relationship. The higher `gdp_per_capita`, the higher `healthy_life_expectancy`.

### Comparing Indonesia's Data with Other Countries'

```
indonesias <- df %>%  
  filter(country == "Indonesia")  
  
others <- df %>%  
  filter(country != "Indonesia")  
  
rows <- c("happiness_score", "gdp_per_capita", "social_support",  
  "healthy_life_expectancy", "freedom_to_make_life_choices", "generosity",  
  "perceptions_of_corruption")  
  
comparison <- data.frame(Index = rows,  
  Indonesia = c(indonesias$happiness_score, indonesias$gdp_per_capita,  
    indonesias$social_support, indonesias$healthy_life_expectancy,  
    indonesias$freedom_to_make_life_choices, indonesias$generosity,  
    indonesias$perceptions_of_corruption),  
  Other = c(mean(others$happiness_score), mean(others$gdp_per_capita),  
    mean(others$social_support), mean(others$healthy_life_expectancy),  
    mean(others$freedom_to_make_life_choices), mean(others$generosity),
```

```

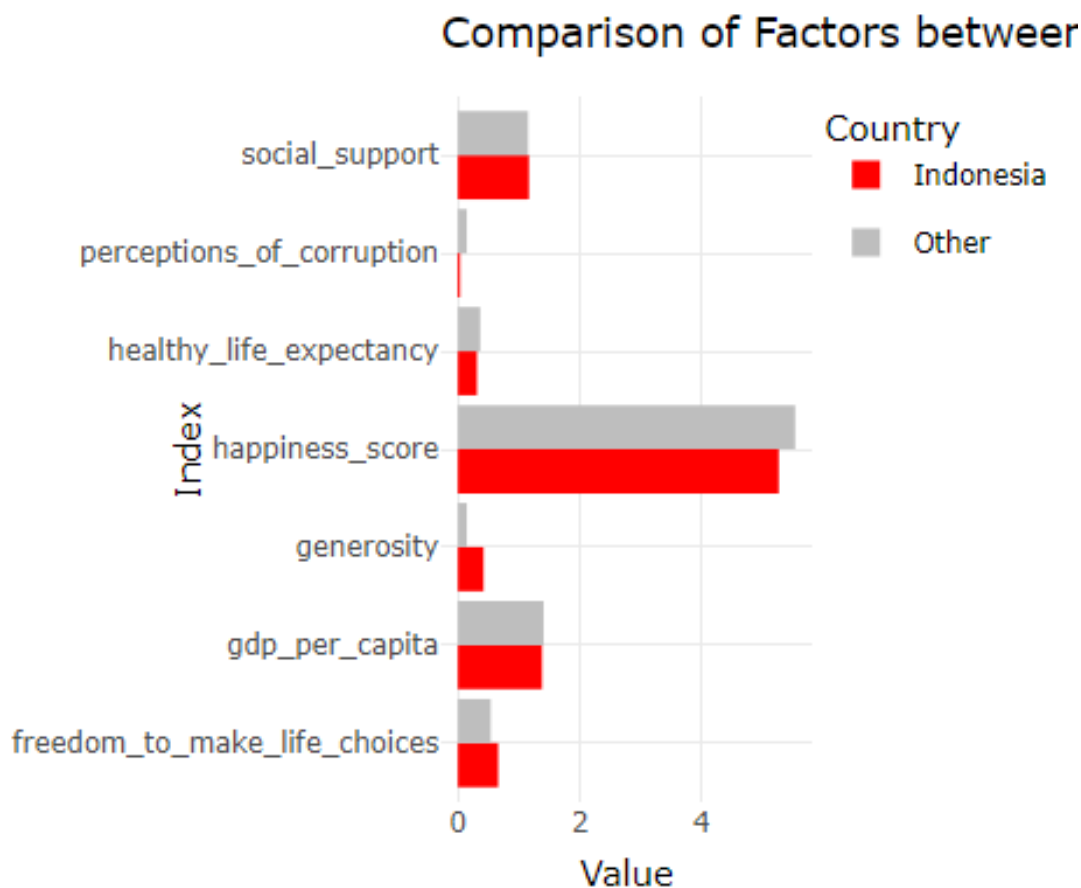
mean(others$perceptions_of_corruption))
)

comparison_long <- tidyr::pivot_longer(comparison, cols = c(Indonesia,
Other),
                                     names_to = "Country", values_to =
"Value")

gg <- ggplot(comparison_long, aes(x = Value, y = Index, fill = Country)) +
  geom_bar(stat = "identity", position = position_dodge()) +
  labs(title = "Comparison of Factors between Indonesia and Other Countries",
       x = "Value", y = "Index") +
  scale_fill_manual(values = c("Indonesia" = "red", "Other" = "grey")) +
  theme_minimal()

ggplotly(gg)

```



Dataframe comparison compares the values of each variable for Indonesia with other countries, which is then illustrated in this grouped bar chart. With this dataframe, we can compare the condition of Indonesia's happiness rate with other countries and the variables that influence it. Based on the dataframe comparison above, Indonesia's happiness\_score is slightly below the average happiness\_score of other countries, as well as variables such as gdp\_per\_capita, healthy\_life\_expectancy, and perceptions\_of\_corruption. Meanwhile,

variables such as social\_support, freedom\_to\_make\_life\_choices, and generosity are slightly above the average for other countries. Thus, Indonesia's overall happiness\_score is 5,277.

## Discussion

Happiness score differs between each region. We can conclude that the top 3 happiest regions are North America and the Australia/New Zealand (ANZ), Western Europe, and Eastern and Central Europe. Meanwhile, the top 5 happiest countries are Finland, Denmark, Iceland, Israel, and the Netherlands.

We did a correlation test between the happiness score variable and each of the factor variables to determine which influences countries' happiness the most. The results show that happiness score is most strongly correlated with social support. This puts emphasis on the importance of social connections and the fact that humans are social beings who thrive on the support of family, friends, and community to enhance their quality of life.

We also explored the impact of financial conditions on happiness. Our correlation test revealed that GDP per capita is highly correlated with healthy life expectancy and social support, with correlations of 84% and 74% respectively. This indicates that higher GDP per capita is associated with better healthcare, education, and infrastructure. Essentially, as GDP per capita increases, people gain greater access to social and healthcare facilities and services.

Lastly, we examined where Indonesia stands compared to other countries. We compared Indonesia's scores for each variable with the average scores of other countries and presented this in the form of a grouped bar chart. Indonesia has a happiness score of 5.28, slightly lower than the global average of 5.54. Indonesia also has slightly lower scores in variables such as GDP per capita, healthy life expectancy, and perception of corruption. However, Indonesia scores higher than the global average in social support, freedom to make life choices, and generosity. In fact, Indonesia has the highest score in terms of generosity. This indicates that while there are areas where improvements can be made, Indonesia excels in social support and generosity.

## Conclusion

In conclusion, these findings underscore the multifaceted nature of happiness and suggest that an approach focusing on social, economic, and health-related factors, is essential for fostering well-being and enhancing happiness globally. Indonesia, in particular, should focus on improving economic conditions and governance to enhance its people's quality of life. By understanding and addressing these key factors, we can work towards creating a happier and more prosperous world for all.

## References

Helliwell, J. F., Layard, R., Sachs, J. D., Aknin, L. B., De Neve, J.-E., & Wang, S. (Eds.). (2023). World Happiness Report 2023 (11th ed.). Sustainable Development Solutions Network.