Initial Analysis: Relationship Between Economic Prosperity and Life Evaluation Expectancy Across Countries

Rafael Juancho San Jose

2025-12-01

Research Question

What is the relationship between economic prosperity and life evaluation across countries from 2021 to 2024?

Hypothesis

I hypothesize that:

1) **Higher log GDP per capita** will be associated with **higher life evaluation**.
2) However, **GDP alone will not fully explain differences in life evaluation**, because:

- Countries with **higher social support** likely experience better health outcomes, even at similar income levels.
- Countries with **higher perceptions of corruption** may have weaker public institutions, potentially reducing health system effectiveness.

Therefore, social and institutional factors (e.g., support networks and corruption perceptions) will partially mediate or modify the relationship between GDP and life evaluation.

Variables

Each variable is measured on a Cantril ladder score on a scale of 0-10.

1) Outcome Variable - Life Evaluation (3-year average)
2) Primary Explanatory Variable - Log GDP per capita
3) Additional Explanatory Variables - social support, healthy life expectancy, freedom, generosity, corruption

Load the dataset

```
# load excel file
df <- read_excel("WHR25_Data_Figure_2.1v3 (1).xlsx")
head(df)
```

```
## # A tibble: 6 x 13
##    Year  Rank `Country name` `Life evaluation (3-year average)` `Lower whisker`
##   <dbl> <dbl> <chr>                                      <dbl>           <dbl>
## 1  2024   147 Afghanistan                                 1.36            1.30
## 2  2023   143 Afghanistan                                 1.72            1.67
## 3  2022   137 Afghanistan                                 1.86            1.80
## 4  2021   146 Afghanistan                                 2.40            2.34
## 5  2020   150 Afghanistan                                 2.52            2.45
## 6  2019   153 Afghanistan                                 2.57            2.51
## # i 8 more variables: `Upper whisker` <dbl>,
## #   `Explained by: Log GDP per capita` <dbl>,
## #   `Explained by: Social support` <dbl>,
## #   `Explained by: Healthy life expectancy` <dbl>,
## #   `Explained by: Freedom to make life choices` <dbl>,
## #   `Explained by: Generosity` <dbl>,
## #   `Explained by: Perceptions of corruption` <dbl>, ...
```

Data Cleaning

```
# rename variables
df <- df %>%
  rename(
    country = "Country name",
    life_eval = "Life evaluation (3-year average)",
    log_gdp = "Explained by: Log GDP per capita",
    social_support = "Explained by: Social support",
    healthy_life = "Explained by: Healthy life expectancy",
    freedom = "Explained by: Freedom to make life choices",
    generosity = "Explained by: Generosity",
    corruption = "Explained by: Perceptions of corruption"
  )
```

```
# filter for years 2021 - 2024
df <- df %>%
  filter(Year >= 2021, Year <= 2024) %>%

  # select only variables used for analysis
  select(Year, country, life_eval, log_gdp, social_support,
         healthy_life, freedom, generosity, corruption)
```

```
# remove NA values
df <- na.omit(df)
head(df)
```

```
## # A tibble: 6 x 9
##    Year country life_eval log_gdp social_support healthy_life freedom generosity
##   <dbl> <chr>       <dbl>   <dbl>          <dbl>        <dbl>   <dbl>      <dbl>
## 1  2024 Afghan~      1.36   0.649          0            0.155   0          0.075
## 2  2023 Afghan~      1.72   0.628          0            0.242   0          0.091
## 3  2022 Afghan~      1.86   0.645          0            0.087   0          0.093
## 4  2021 Afghan~      2.40   0.758          0            0.289   0          0.089
## 5  2024 Albania      5.41   1.37           1.12         0.696   0.841      0.103
## 6  2023 Albania      5.30   1.44           0.924        0.638   0.69       0.138
## # i 1 more variable: corruption <dbl>
```

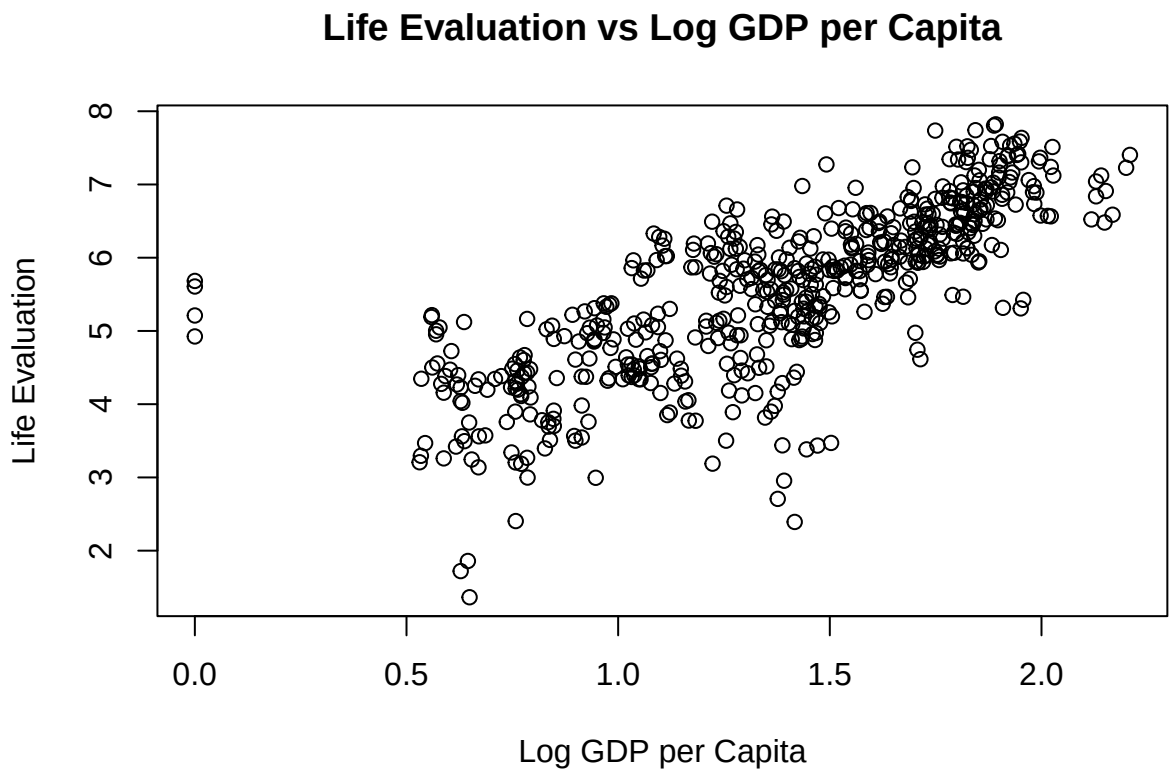
Summary Statistics

```
summary(df)
```

```
##       Year        country            life_eval        log_gdp
##  Min.   :2021   Length:566         Min.   :1.364   Min.   :0.000
##  1st Qu.:2021   Class :character   1st Qu.:4.724   1st Qu.:1.078
##  Median :2023   Mode  :character   Median :5.766   Median :1.425
##  Mean   :2022                      Mean   :5.553   Mean   :1.380
##  3rd Qu.:2024                      3rd Qu.:6.414   3rd Qu.:1.728
##  Max.   :2024                      Max.   :7.821   Max.   :2.209
##  social_support   healthy_life       freedom         generosity
##  Min.   :0.000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.888   1st Qu.:0.3723   1st Qu.:0.4940   1st Qu.:0.0870
##  Median :1.179   Median :0.5265   Median :0.6205   Median :0.1300
##  Mean   :1.130   Mean   :0.5082   Mean   :0.6074   Mean   :0.1390
##  3rd Qu.:1.392   3rd Qu.:0.6478   3rd Qu.:0.7278   3rd Qu.:0.1845
##  Max.   :1.840   Max.   :0.9480   Max.   :1.0180   Max.   :0.4680
##    corruption
##  Min.   :0.00000
##  1st Qu.:0.06425
##  Median :0.11700
##  Mean   :0.15037
##  3rd Qu.:0.19050
##  Max.   :0.58700
```

Plots of Key Dependent and Explanatory Variables

```
# scatter plot of life_eval and log_gdp
plot(df$log_gdp, df$life_eval,
     xlab = "Log GDP per Capita",
     ylab = "Life Evaluation",
     main = "Life Evaluation vs Log GDP per Capita")
```

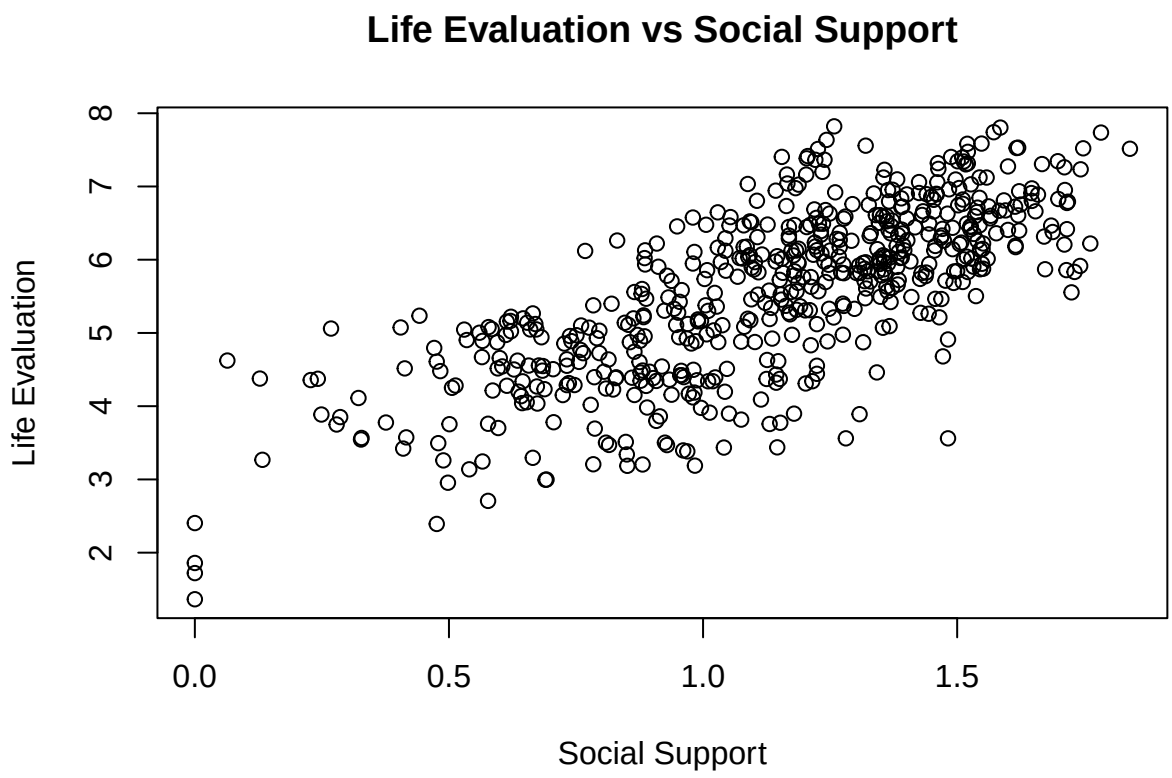



```
# correlation coefficient of life evaluation and log gdp per capita
cor(df$life_eval, df$log_gdp)
```

```
## [1] 0.7649693
```

Life Evaluation is strongly positively correlated with Log GDP per capita, indicating that countries with higher GDP per capita tend to have higher life evaluation scores. Correlation coefficient is around 0.76.

```
# scatter plot of life_eval and social_support
plot(df$social_support, df$life_eval,
     xlab = "Social Support",
     ylab = "Life Evaluation",
     main = "Life Evaluation vs Social Support")
```

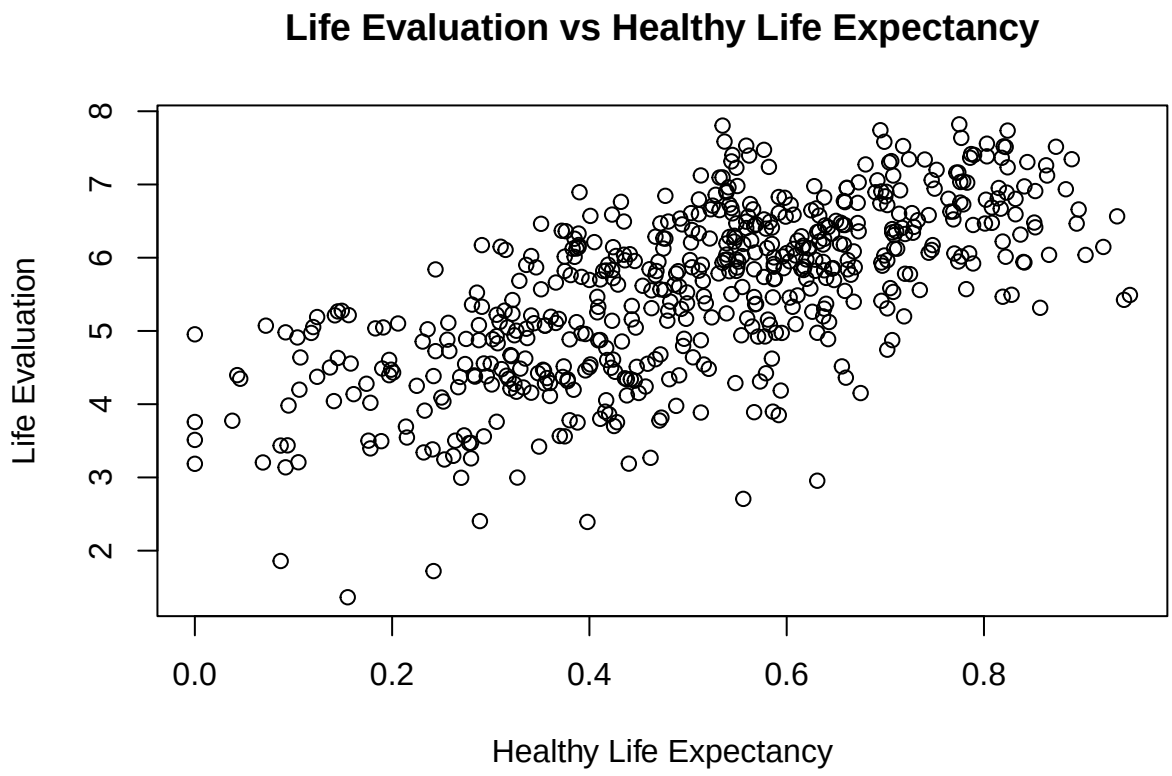



```
# correlation coefficient of life evaluation and social support
cor(df$life_eval, df$social_support)
```

```
## [1] 0.736975
```

Life Evaluation is strongly positively correlated with social support, indicating that countries with higher social support tend to have higher life evaluation scores. Correlation coefficient is around 0.74.

```
# scatter plot of life_eval and healthy_life
plot(df$healthy_life, df$life_eval,
     xlab = "Healthy Life Expectancy",
     ylab = "Life Evaluation",
     main = "Life Evaluation vs Healthy Life Expectancy")
```

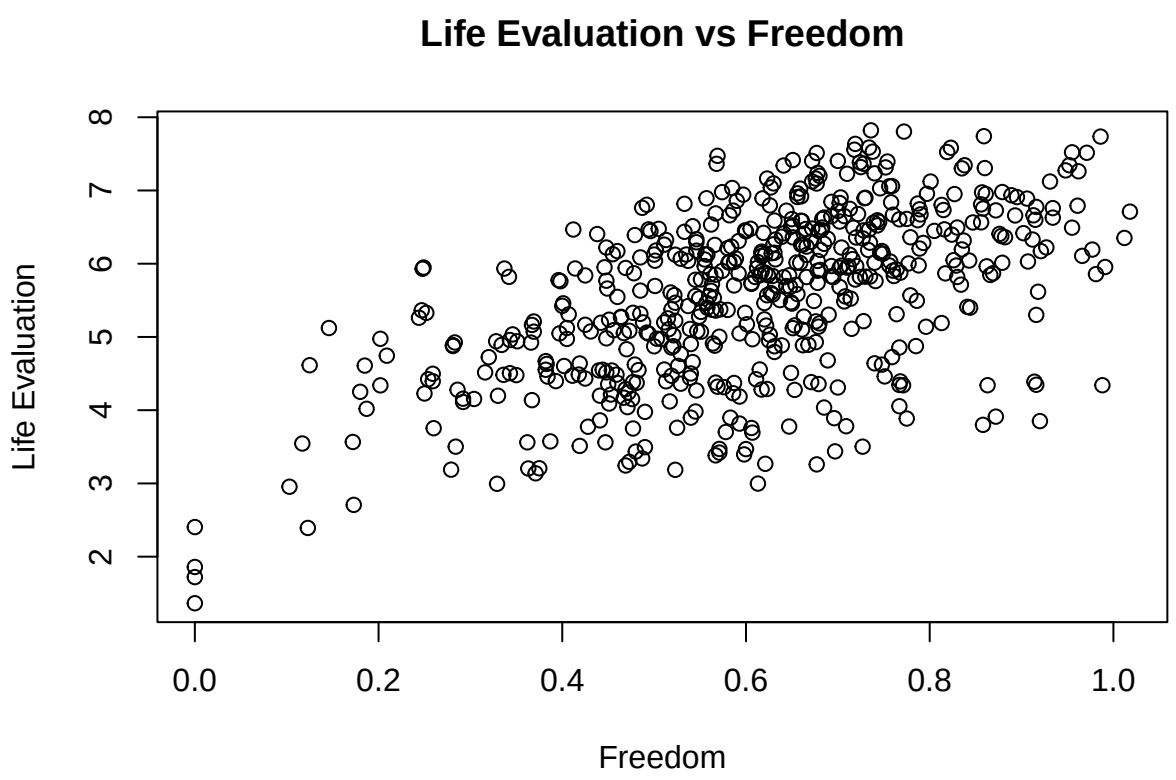



```
# correlation coefficient of life evaluation and healthy life
cor(df$life_eval, df$healthy_life)
```

```
## [1] 0.6668539
```

Life Evaluation is moderately positively correlated with healthy life expectancy, indicating that countries with higher healthy life expectancy tend to have higher life evaluation scores. Correlation coefficient is around 0.67. It's slightly weaker than the correlations for log_gdp (~0.76) and social support (~0.74) but still meaningful.

```
# scatter plot of life_eval and freedom
plot(df$freedom, df$life_eval,
     xlab = "Freedom",
     ylab = "Life Evaluation",
     main = "Life Evaluation vs Freedom")
```

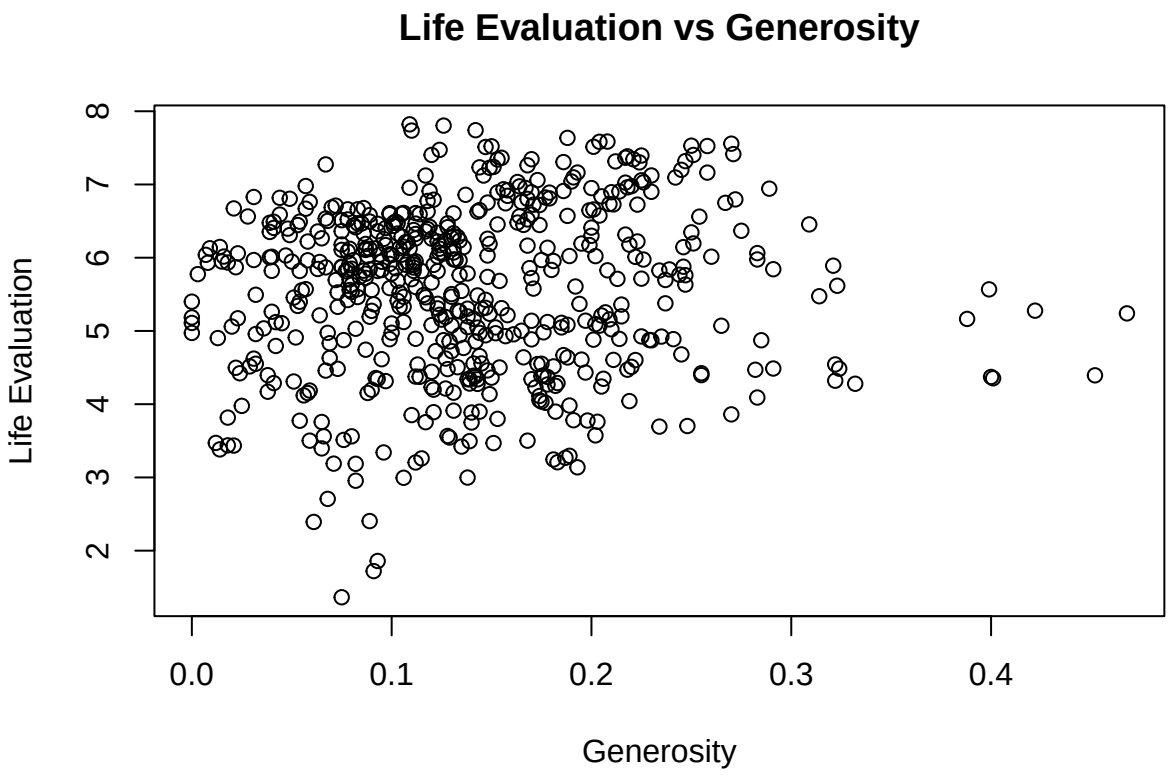



```
# correlation coefficient of life evaluation and freedom
cor(df$life_eval, df$freedom)
```

```
## [1] 0.5596448
```

Life Evaluation is moderately positively correlated with freedom to make life choices, indicating that countries where people report more freedom tend to have higher life evaluation scores, although this linear association is weaker than that of GDP, social support, or healthy life expectancy. Correlation coefficient is around 0.56.

```
# scatter plot of life_eval and generosity
plot(df$generosity, df$life_eval,
     xlab = "Generosity",
     ylab = "Life Evaluation",
     main = "Life Evaluation vs Generosity")
```

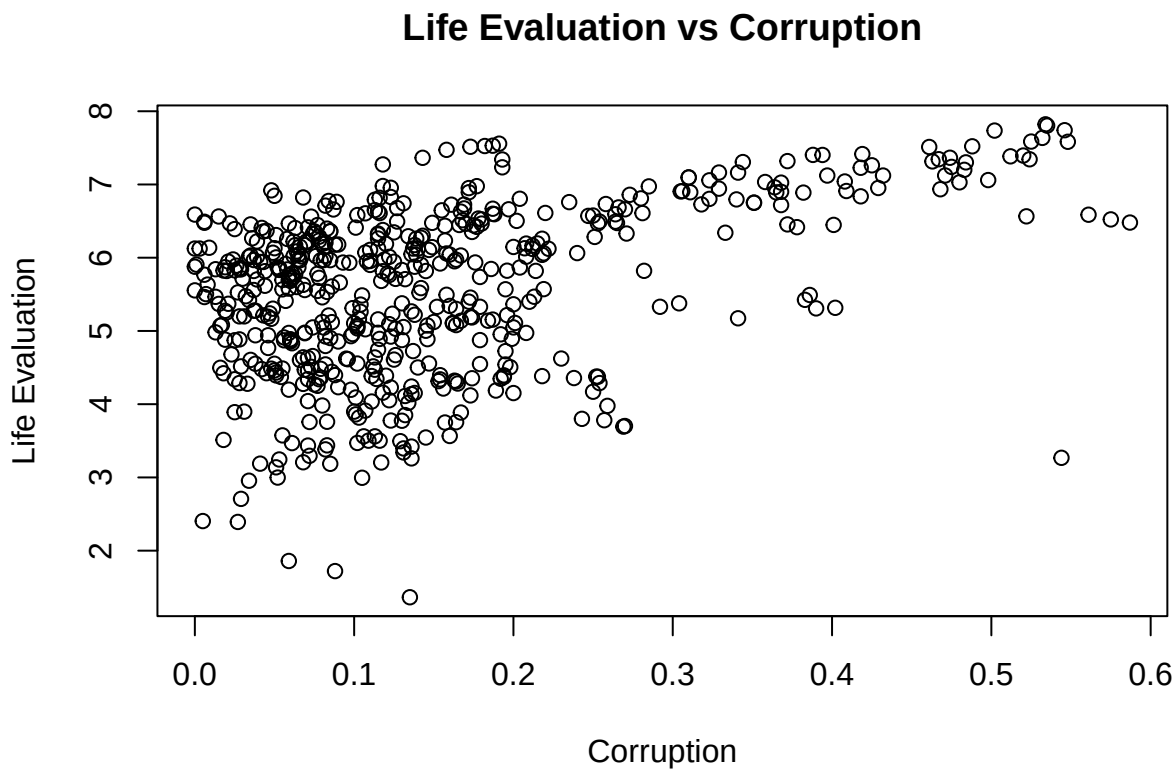



```
# correlation coefficient of life evaluation and generosity
cor(df$life_eval, df$generosity)
```

```
## [1] 0.06648802
```

Life Evaluation shows almost no linear association with Generosity , suggesting that differences in generosity across countries do not have a strong linear association to differences in life evaluation scores.Correlation coefficient is around 0.07.

```
# scatter plot of life_eval and corruption
plot(df$corruption, df$life_eval,
     xlab = "Corruption",
     ylab = "Life Evaluation",
     main = "Life Evaluation vs Corruption")
```




```
# correlation coefficient of life evaluation and corruption
cor(df$life_eval, df$corruption)
```

```
## [1] 0.437471
```

Life Evaluation is moderately positively correlated with Perceptions of Corruption, indicating that countries with lower perceived corruption tend to report higher life evaluation scores. Correlation coefficient is around 0.44.

Main Regression

```
# multiple linear regression
fit <- lm(life_eval ~ log_gdp + social_support + healthy_life + freedom + corruption, data = df)
summary(fit)
```

```
##
## Call:
## lm(formula = life_eval ~ log_gdp + social_support + healthy_life +
##     freedom + corruption, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.81793 -0.30255  0.07394  0.36629  1.39565
```

```
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     1.80261    0.10099  17.850  < 2e-16 ***
## log_gdp         0.98060    0.09424  10.405  < 2e-16 ***
## social_support  1.01082    0.09990  10.118  < 2e-16 ***
## healthy_life    0.83965    0.17383   4.830 1.76e-06 ***
## freedom         1.10011    0.16569   6.639 7.46e-11 ***
## corruption      1.06201    0.22200   4.784 2.20e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5667 on 560 degrees of freedom
## Multiple R-squared:  0.7558, Adjusted R-squared:  0.7536
## F-statistic: 346.6 on 5 and 560 DF,  p-value: < 2.2e-16
```

We ran a multiple linear regression with Life Evaluation as the dependent variable and five explanatory variables: Log GDP per capita, Social Support, Healthy Life Expectancy, Freedom to make life choices, and Perceptions of Corruption.

All five predictors are positively associated with Life Evaluation. Specifically, holding other factors constant:

- A 1-unit increase in Log GDP per capita is associated with an increase of approximately 0.98 points in Life Evaluation.
- A 1-unit increase in Social Support increases Life Evaluation by ~1.01 points.
- Healthy Life Expectancy contributes ~0.84 points per unit.
- Freedom to make life choices adds ~1.10 points.
- Higher Perceptions of Corruption contribute ~1.06 points (note: higher values indicate lower corruption).

All coefficients are highly statistically significant ($p < 0.001$), indicating that these variables reliably predict differences in Life Evaluation across countries.

These results support our hypothesis that GDP alone does not fully explain Life Evaluation; social support, freedom, and corruption also play meaningful roles. The positive coefficients align with the expectation that higher income, stronger social networks, greater freedom, and lower corruption are associated with higher reported life evaluation.