

PROJECT Planning

| | |
|----------------------|--------------------------------------|
| Class Section | STAT-614-001.2025S Statistic Methods |
| GROUP MEMBERS | Qintao Jia, Anna Hyunjung Kim |
| TITLE | World Happiness Report |

DATA set (detailed description of your data set) with list of variables you will use (mentioning units of measurements)

Data source: Sazidthe1. (2023). World Happiness Report (till 2023) [Data set]. Kaggle.

<https://www.kaggle.com/datasets/sazidthe1/global-happiness-scores-and-factors>

Data Originally from <https://worldhappiness.report/>

1. The World Happiness Report (2020)

a. 153 unique observations

b. 9 columns

- i. country
- ii. region
- iii. happiness_score
- iv. gdp_per_capita
- v. social_support
- vi. healthy_life_expectancy
- vii. freedom_to_make_life_choice
- viii. generosity
- ix. perceptions_of_corruption

2. The World Happiness Report (2015)

a. 158 unique observations

b. 9 columns

- i. country
- ii. region
- iii. happiness_score
- iv. gdp_per_capita
- v. social_support
- vi. healthy_life_expectancy
- vii. freedom_to_make_life_choice
- viii. generosity
- ix. perceptions_of_corruption

| Variable | Description | Type | Unit |
|------------------------------|--|---------------------------|------------------------|
| country | Name of the country | Categorical | N/A |
| region | Region the country belongs to | Categorical | N/A |
| happiness_score | Average life evaluation score based on the Cantril ladder (0–10 scale) | Quantitative | Score (0–10) |
| gdp_per_capita | Log of GDP per capita based on Purchasing Power Parity (PPP) in constant 2021 international dollars | Quantitative | Log(USD, PPP-adjusted) |
| social_support | National average of the binary response (0 = no, 1 = yes) to the question: “Do you have someone to rely on in times of trouble?” | Quantitative (proportion) | Proportion (0–1) |
| healthy_life_expectancy | Life expectancy at birth adjusted for health quality | Quantitative | Years |
| freedom_to_make_life_choices | National average of responses to: “Are you satisfied with your freedom to choose what you do with your life?” (0 = no, 1 = yes) | Quantitative (proportion) | Proportion (0–1) |
| generosity | Residual from regressing donation behavior on GDP per capita; measures prosocial behavior | Quantitative | Index (no unit) |
| perceptions_of_corruption | National average of responses to corruption-related questions (0 = not widespread, 1 = widespread) | Quantitative (proportion) | Proportion (0–1) |

| Research questions | |
|--------------------|--|
| Q1 | Do the happiness scores of people on different continents in 2015 differ significantly |
| Q2 | Have the happiness scores of people in various countries changed between 2015 and 2020 (before and after the COVID-19) |
| Q3 | How do the different explanatory variables influence the happiness score in 2015 |

| TENTATIVE Plan of Action (techniques to be used to address each question listed above) | |
|--|---------------------|
| <p>(a) We will use descriptive statistics and boxplots to explore the data. These summaries will help us understand the distribution, variability, and outliers in the data.</p> <p>(b) We will use the Shapiro-Wilk test and Q-Q plots to assess whether the variables are normally distributed (Normality). Moreover, we will use Levene’s test for ANOVA to verify homogeneity of variances. Lastly, Residual vs. fitted plots and Q-Q plots of residuals will be used for linearity and normality of residuals.</p> <p>(c) Statistical procedures.</p> | |
| Q1 | ANOVA |
| Q2 | Pair t-test |
| Q3 | Multiple regression |

| Group members | Responsibilities |
|-------------------|----------------------------------|
| Qintao Jia | Project plan |
| Anna Hyunjung Kim | Search for dataset, Project plan |

Box 2.2: Explaining the variables in Table 2.1

GDP per capita is in terms of Purchasing Power Parity (PPP) adjusted to constant 2021 international dollars, taken from the World Development Indicators (WDI) by the World Bank (version 34, last updated on 28 October 2024). See the online statistical appendix for more details. GDP data for 2024 are not yet available, so we extend the GDP time series from 2023 to 2024 using country-specific forecasts of real GDP growth from the OECD Economic Outlook No. 116 (Edition 2024/2) or, if missing, from the World Bank’s Global Economic Prospects (last updated: 11 June 2024), after adjustment for population growth. The equation uses the natural log of GDP per capita as this form fits the data significantly better than GDP per capita.

The time series for **healthy life expectancy** at birth are constructed based on data from the World Health Organization (WHO) Global Health Observatory data repository, with data available up to 2021 (last updated: 2 August 2024). To match this report’s sample period (2005–2024), interpolation (when necessary) and extrapolation are used. See the online statistical appendix for more details.

Social support is the national average of the binary responses (0=no, 1=yes) to the Gallup World Poll (GWP) question “If you were in trouble, do you have relatives or friends you can count on to help you whenever you need them, or not?”

Freedom to make life choices is the national average of binary responses to the GWP question “Are you satisfied or dissatisfied with your freedom to choose what you do with your life?”

Generosity is the residual from regressing the national average of GWP responses to the donation question “Have you donated money to a charity in the past month?” on log GDP per capita.

Perceptions of corruption is the average of binary answers to two GWP questions: “Is corruption widespread throughout the government or not?” and “Is corruption widespread within businesses or not?” Where data for government corruption are missing, the perception of business corruption is used as the overall corruption-perception measure.

Positive affect is defined as the average of previous-day affect measures for laughter, enjoyment, and doing interesting things. The inclusion of doing interesting things (first added for World Happiness Report 2022) gives us three components in each of positive and negative affect, and slightly improves the equation fit in column 4. The general form for the affect questions is: “Did you experience the following feelings during a lot of the day yesterday?” See the online statistical appendix for more details.

Negative affect is defined as the average of previous-day affect measures for worry, sadness, and anger.