

# Final Project Description

## Background and Introduction

The final project for the course will require you to complete some exploratory analyses based on data from the World Happiness Report, which has been uploaded to Kaggle.

<https://www.kaggle.com/unsdsn/world-happiness/data>

I have attached the dataset to the project description on Crowdmark so that you do not have to register for an account. The data on Kaggle is comprised of five spreadsheets in different formats. The data for each country includes the country name, the region, and six characteristics that contribute the most to the predicted happiness score for a country (measured on an arbitrary scale).

The dataset I've attached below contains data on countries who reported at least some data for all five years. I've also recorded the regions so that they are easier to analyze. The variables for this project are summarized in the table below:

Variable	Description
Country	Country name
Region	Region of the world
GDPperCap	Economic Production score
TrustGov	Trust in Government score
Family	Family Support score
Freedom	Perceived Freedom score
HealthLifeExp	Perceived Health and Life Expectancy score
Generosity	Perceived Generosity score
Year	Year of the survey

## Objectives and evaluation

The project requires you to complete three tasks, detailed below. You should prepare a single report containing your answers to all tasks. Include the code for each task in your report, for reproducibility purposes. You may include the code as code chunks where the analyses are taking place or, if you prefer, you may include it at the end (although the code should be clearly commented so that it is clear which task each block of code corresponds to).

The completion of each task is worth 25 points. The quality of presentation will also be worth 25 points, i.e. clarity of explanation, plots, tables, and code.

The length of the projects will vary, depending on the number and formatting of figures and tables and the conciseness of the writing. Rather than focusing on the number of pages, I encourage students to focus on completing each task (and subtask) below to the best of their ability in the clearest and most efficient manner.

## Tasks to complete

### Task 1: Family, Generosity and Freedom by Region in 2019

The first task is to provide some exploratory data analyses and describe the distributions of the six scores **by region** for **only the year 2019**.

- (a) Using the appropriate plots and numerical summaries, evaluate the strength of the association of each of the six scores with the region variable.
- (b) Does the association of between score and region vary amongst the scores or do all the scores seem to seem to show the same pattern? Explain briefly the reasons behind your assessment.

### Task 2: Predicted Happiness score

The six World Happiness Report scores for each country all contribute to the predicted happiness for each country based on these six factors. One can obtain a total overall happiness score for each **country** for each *year* by taking the sum total of the six scores (Economic Production, Trust in Government, Family Support, Freedom, Generosity, and Health+Life Expectancy) for that country in that year.

- (a) Compute the happiness score for each country for each year.
- (b) Compute the minimum and maximum happiness score over all countries for each region in each year. In other words, you should have a minimum and maximum happiness score for each region for each year. Plot how the minimum and maximum happiness score for each region varies over time, i.e. with year on the x-axis and lines for the minimum and maximum for each region on the y-axis. **Hint:** You will want to have multiple panels in your plot, either one each for the minimum and maximum OR one for each region. Do the minimum and/or maximum happiness scores change much over time? Explain your answer.

### Task 3: Happiest((well, predicted happiest) countries over time

Using the country happiness scores from Task 2:

- (a) Create a table with the 10 **countries** who have the highest average happiness scores over the five years of the data.
- (b) Create a table with the 10 **countries** who have the largest positive change in happiness score from 2015 to 2019, i.e. the 10 countries with the largest positive value of the difference between their 2019 happiness score and their 2015 happiness score.