

Tutorial 9 Worksheet AY 22/23 Sem 1

DSA2101

The dataset for this week is derived from the World Happiness Report 2022. You can download the excel sheet `happiness_2022_v2.xls` from Canvas. It contains the happiness scores for countries in the year 2022. Happiness scores are based on a survey of the residents of the country.

These are the columns in the dataset:

1. **RANK**: the happiness ranking of the country. 1 is the happiest country.
2. **Country**: the name of the country.
3. **Happiness**: the happiness score of the country.
4. Columns 4 - 9: Using a regression model, the happiness score of each country is broken down into 6 components:
 - **Dystopia + residual** - an unexplained component of the happiness.
 - **GDP** - Gross Domestic Product
 - **Social support**
 - **Healthy Life Expectancy**
 - **Freedom to make Life Choices**
 - **Generosity**
 - **Perceptions of Corruption**

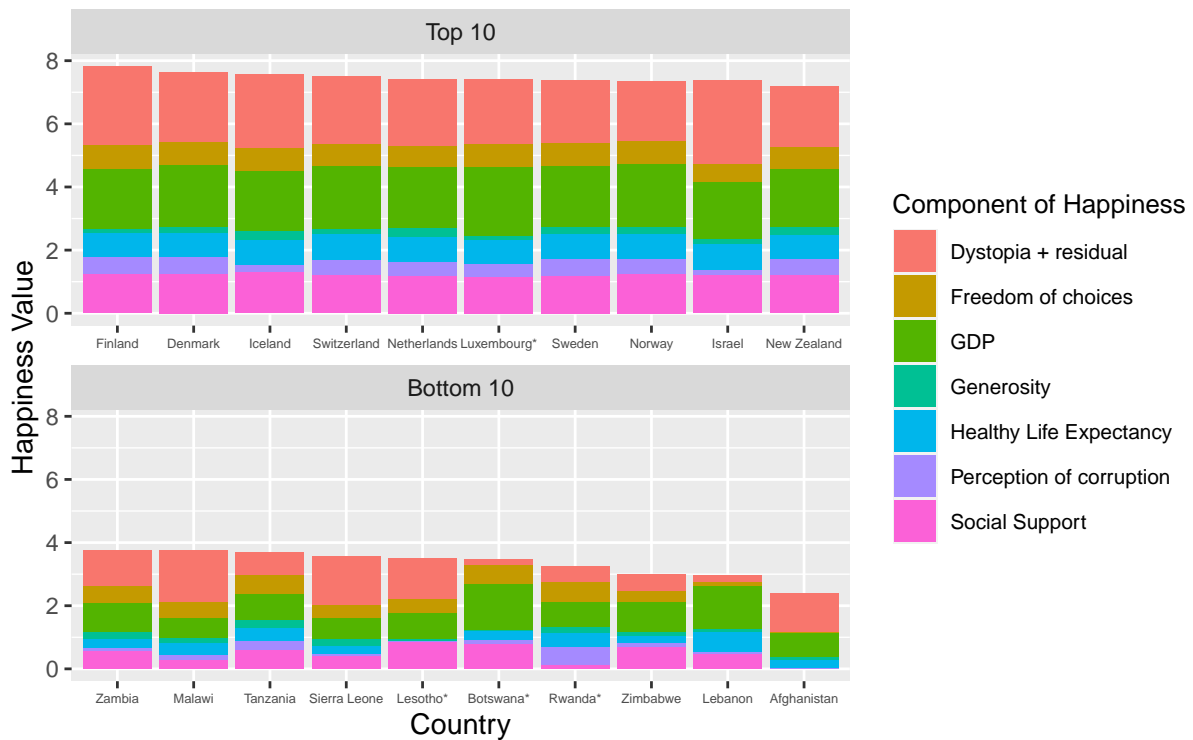
The idea is to try to attribute happiness to these 6 components using a model.

5. **Region**: the geographical region the country belongs to.

Question 1

Re-create the following plot, which tries to compare how the individual components compare between the happiest and least happiest countries:

Components of happiness for top 10 and bottom 10 ranked countries



Question 2

Re-create the following boxplot, which compares how happiness scores vary from region to region.

Happiness scores by region

