Section B – Global happiness index (10 marks)

The World Happiness Report ranks countries on the perceived happiness of their citizens. The happiness scores are based partly on data from the Gallup World Poll which is a nationally representative annual survey of each country's population aged 15 and over.

There are five domains of happiness: social support, healthy life expectancy at birth, freedom to make life choices, perceptions of corruption, and generosity. Another strong predictor of happiness is a country's GDP per capita (Global Domestic Product per one hundred thousand population).

In this section you will explore, visualise, and interpret global happiness scores over the period 2015—2019. The supplied datasets are:

A2B happiness_2015.csv
A2B happiness_2016.csv
A2B happiness_2017.csv
A2B happiness_2018.csv
A2B happiness_2019.csv

Question 1

Create a long format dataset for 2015—2019 containing the following variables:

- Year
- Country
- Happiness_rank
- Happiness_score
- GDP_per_capita
- Social_support
- Healthy_life_expectancy
- Freedom_to_make_life_choices
- Perceptions_of_corruption
- Generosity

a) You will need to harmonise the variable names since they are named differently in the older datasets (2015—2017) than in the newer ones (2018—2019). For example, the variable 'Family' is the same as 'Social support'.
b) You will need to edit some country names. For example, "Macedonia" is now called "North Macedon". Check that each country only has one observation per year.
c) Briefly document your decisions for (a) and (b) using code snippets and/or dot-points and tables. Show a frequency table for the variable Year for the long dataset.

[3 marks]

Question 2

a) List the happiness ranks and happiness scores for Australia and the countries that were in the top 5 or the bottom 5 in 2019. **[1 mark]**

b) Create graphs for the change in happiness index over time for the two groupings identified in part (a): the top 5 plus Australia and then the bottom 5. Let SAS choose the natural scale for the vertical axis.

Now show all 11 countries on the same graph. How do we interpret what these two types of plots are telling us in terms of the scale? **[2 marks]**

c) Create a scatterplot of life expectancy by GDP per capita for the year 2019. How do we interpret the position of Saudi Arabia relative to Hong Kong in the scatterplot?

[1 mark]

d) How well does economy (measured in terms of GDP per capita) correlate with the other four measures in the happiness index? **[1 mark]**

e) Use the geographical region variable from the 2015 dataset and merge this information into the stacked dataset. Show a frequency table of region by year. **[1 mark]**

f) Create a single plot that shows the distribution of happiness score by region. Only include regions in Asia, Europe, and Africa and make sure the regions appear grouped within these three super regions. **[1 mark]**