R for Economics and Social Science Research

WORKSHOPS

2025-03-31

Data Set

The World Happiness Report Up to 2023 dataset offers a comprehensive and up-to-date examination of happiness metrics and the factors influencing well-being on a global scale. This dataset is designed to provide valuable insights for policymakers, researchers, and individuals interested in understanding the dynamics of happiness and well-being worldwide.

These datasets include key metrics related to global happiness and well-being, such as country names, regions, happiness scores, GDP per capita, social support, healthy life expectancy, freedom to make life choices, generosity, and perceptions of corruption. These metrics offer insights into the happiness and socio-economic conditions of various countries and regions, making it a valuable resource for analyzing and understanding well-being on a global scale.

(https://www.kaggle.com/datasets/sazidthe1/global-happiness-scores-and-factors)

WORKSHOP 1

1. Load the "*World happiness data 2018-2023.xlsx*" file into the R workspace.
2. Prepare a summary table of the mean, median, and standard deviation of one particular variable of interest to you (say, *happiness score* or *healthy life expectancy*) by country and year. Provide an interesting story based on the table.
3. Using the 2022 and 2023 data sets, generate the mean, median, and standard deviation of all variables for all countries in Asia. What interesting facts emerge from the results?
4. a) Using the 2023 data, calculate the mean of the "*corruption index*" by region.
   b) Add these means to the original data frame using either the merge() or the left_join() function. [Read the documentation of these functions to learn their syntax.].
   c) Classify each country as **low**, if the country's corruption index is below the regional mean, or **high** if the country's corruption index is above the regional mean. Create a new variable which contains this information, and label it as *corrupt_level*.
   d) Create a cross tabulation of *corrupt_level* and *region*.