World Inequality and Happiness Summary

# Project Topic

I will be analyzing data from the World Inequality Database (WID) and the World Happiness Report (WHR). I will be attempting to find and summarize relationships between economic variables and happiness scores for 52 countries in the year 2015.

# Objective

The objective of my analysis is to show the relationships between economic variables in countries and the relationships between these variables and the countries’ overall happiness scores.

# Intended Audience

The intended audience of my analysis is anyone who is interested in the relationships between economic facts of countries and the happiness of their citizens. Those involved in government may be especially interested in the analysis as it may help them to inform decision-making with respect to how to best improve the happiness of their constituency.

# Previous Research

I have not conducted any previous research and have no current knowledge of my topic.

# Project Design

The analysis starts by selecting a subset of economic variables from the WID. The variables selected are as follows:

* LCU per USD: This is the market exchange rate of the country’s local currency units (LCU) to US dollars (USD). This is necessary to convert economic variables from the country’s local currency into USD so that they can be compared between countries.
* Net national income per capita
* Average primary income: This refers to income that is earned through the production of goods and services. It includes the income generated by individuals or entities from factors of production such as labor, capital, and land.
  + Household & non-profits
  + Corporate
  + Government
* Average secondary income: This, also known as transfer income or redistributive income, consists of transfers of money or goods between individuals, households, or countries. These transfers are often aimed at redistributing income or providing financial assistance.
  + Household & non-profits
  + Corporate
  + Government
* Average household wealth
  + All households
  + Bottom 99% of households
  + Top 1% of households
* Social security benefits
* Share of net national income
  + Labor
  + Capital

From the WHR only one variable is selected: the happiness ladder score. This is a value from 0-10, with 0 representing the worst possible life and 10 representing the best possible life. Citizens of each country are asked to subjectively rate their lives on this scale and the country’s ladder score is the average of these ratings in each country.

For the WID data, not all variables are defined for every country for every year, and WHRs are only available from 2012-2023. So the first step is to select a year in this range which maximizes how many countries have all the selected variables defined. After this is done, it is seen that 2015 is the best such year, giving 52 countries.

The data then needs to be collected from the WID CSVs and combined with the happiness score from the WHR for 2015. Once this is done, economic variables need to be normalized into one currency (as opposed to the countries’ local currencies). The US dollar is selected for this purpose, and the economic variables for each country are updated accordingly.

Once this is done, there is no more data cleaning necessary. There are no null values nor duplicate records, and all values are of the appropriate datatype. A series of scatter plots can then be produced to show the correlations between economic variables and happiness.

# Analysis

I’ll be primarily constructing scatter plots to show the correlations between variables across countries.

# Data Collection

The data is collected from the [World Inequality Database](https://wid.world/data/) and the [World Happiness Report](https://worldhappiness.report/data/) for 2015. For the WID, all indicators across all countries and all years were selected. For the WHR, the Excel spreadsheet [“Chapter 2 Online Data Expanded with Trust and Governance”](https://s3.amazonaws.com/happiness-report/2015/Chapter2OnlineData_Expanded-with-Trust-and-Governance.xlsx) was used, as it contains the happiness ladder scores for each country in the worksheet named “Data for Figure2.2”.

# Intended Outcome

The intended outcome of this analysis is a deeper understanding of the relationships between economic variables in countries and countries’ happiness scores.