Preliminary Analysis

**Name:** Jordy Dickinson

**Project Title:** World Inequality and Happiness

**Project Objective:** Analyze the relationships between countries’ economic variables and the happiness of their citizens.

**Data Collection Source:** Economic data comes from the [World Inequality Database](https://wid.world/data/) downloaded for all indicators, all countries/regions, and all years. Happiness data comes from the [World Happiness Report for 2015](https://s3.amazonaws.com/happiness-report/2015/Chapter2OnlineData_Expanded-with-Trust-and-Governance.xlsx). The year 2015 was chosen because, after selecting a subset of the available economic variables, it was found that the year 2015 had the most countries without null values for these variables, after restricting the range of years to 2012-2023 (the years for which World Happiness Reports are published).

**Methodology:** Using the [codes dictionary](https://wid.world/codes-dictionary/) of the World Inequality Database, a subset of the available variables was chosen. Using Python, the year for which the most countries had no null values for these variables between the years 2012-2023 was determined (2015), at 52 countries in total. The relevant values for these variables were collected from the World Inequality Database CSV files, and monetary values were converted to USD (using the market exchange rate which is also available in this dataset). This data was combined with the happiness ladder score from the World Happiness Report data for 2015. The resultant subset of these two datasets had no null or missing values, no type conversions were necessary, and there were no typos.

**Findings:**

1. **Data Size:** The data contains 16 columns and 52 rows.
2. **Missing Values:** The World Inequality Database often had missing values for country/year/variable combinations. However, a subset of this data was selected for which there were no missing values, giving 52 countries to analyze. There were no missing values in the relevant World Happiness Report data.
3. **Duplicates:** There were no duplicates in the data.
4. **Dates:** All data was for the year 2015.
5. **Additional Findings:** There were no additional findings.