

M04 - R Programming Language

● Live Lesson Challenge

R

R-evolutionize Your Data Analysis

In this challenge, you will apply your data manipulation, visualization, and basic statistical analysis skills to the World Happiness Report 2023 dataset. You will clean and prepare the data, create visualizations, and perform a basic statistical analysis to compare happiness scores across high and low GDP countries.

Database: [World Happiness Report 2023](#)

The World Happiness Report uses data from the Gallup World Poll to estimate happiness scores and rankings across 6 key factors: economic production, social support, life expectancy, freedom, absence of corruption, and generosity. These factors contribute to making life evaluations higher or lower in each country compared to Dystopia, a hypothetical baseline country with the world's lowest national averages for each factor.

Challenge Tasks:

1. Data Exploration and Cleaning :

- **Explore the dataset:** Load the dataset and check its structure, summary statistics, and the first few rows.
- **Create a GDP category:** Classify countries into **High GDP** or **Low GDP** based on whether their `Logged.GDP.per.capita` is above or below the median value.
 - **Hint:** Use the `median()` function to find the median GDP, and `ifelse()` to categorize the countries.
- **Clean the data:** Remove any rows where the happiness score `Ladder.score` is missing

2. Data Summarization

- **Calculate average happiness scores:** Group the dataset by **GDP category** (high vs. low GDP) and calculate the average happiness score `Ladder.score` for each group.

3. Data Visualization

- **Create a box plot:** Create a box plot that compares the happiness scores `Ladder.score` between high and low GDP countries. Use `ggplot2` to create the plot.

4. Statistical Analysis

- **Perform a t-test:** Perform a t-test to compare the average happiness scores between high and low GDP countries. Interpret the result briefly (focus on the p-value).
-