Applied Session 2
Einstein Intro to R

Angel Garcia de la Garza

EINSTEIN
Albert Einstein College of Medicine

Montefiore
nspired Medicine

GDP Dataset

- The dataset contains the GDP (USD), population, and percentage of females for each country from 1960 to 2022.

- Six variables:
    - country: Country Name
    - country_code: Abbreviation
    - year
    - gdp: GDP of country in USD
    - Population: population of country
    - prop_female: % of females in population

Goals

- We want to explore the relationship between GDP and population and the number of medals a country wins at the Olympics.

- We need to merge the gdp dataset with the Olympics dataset

Merging Issues

- The names of the countries in both the Olympics and GDP dataset do not all match each other.

- We first need to identify which countries aren't matching

- This is an example of a common real-life data cleaning

Questions we want to answer:

Example 1: Using the 2000 Summer Olympics as an example, is GDP related to the number of gold medals each country won at the Olympics?

We need to:

1. Filter the data to include only the information from the 2000 Olympics.
2. Calculate the number of gold medals won by each country.
3. Determine the ranking of each country based on both medal count and GDP.
4. We can do this using the "rank()" function

Additional Questions:

- Is a country's GDP proportional to the number of gold medals won? Taking the 2000 Summer Olympics as an example, is GDP related to the medal count (3 gold, 2 silver, 1 bronze) each country achieved?

- Instead of using the 2000 Olympics as an example, could you provide a historical average?

- Using the 2000 Summer Olympics as an example, is GDP per capita related to the number of gold medals each country won at the Olympics?