

Data Wrangling

Thiyanga Talagala

Load the `gapminder` dataset and the `tidyverse` and `magrittr` packages.

Exercises

1. Filter all rows for “Sweden”.
2. Filter all rows where `lifeExp` is less than or equal to 30.
3. Filter all rows that have a missing value for `year`.
4. Filter all countries that had population over 100000 in 1960 or earlier.
5. Count the number of countries with life expectancy greater than 30 in 1952.
6. Calculate the mean life expectancy for each year and continent.
7. Get the maximum and minimum of GDP per capita for all continents in a “wide” format.
8. Get the maximum and minimum of GDP per capita for all continents in a “long” format.
9. What was the population of the United States in 1952 and 2007.
10. Subset the `gapminder` data to extract rows where `lifeExp` is greater than or equal 80. Retain only the columns `country`, `year`, and `lifeExp`. Sort the results from largest to smallest based on `lifeExp`.
11. Calculate the total GDP in billions of dollars, extract the results for the year 2002, and sort the rows so that the total GDP is in decreasing order.

Help: `gpd = gdpPercap * pop`

12. Calculate the average life expectancy by continent in 2002.
13. Which countries and which years had the worst five GDP per capita measurements?
14. What was the mean life expectancy across all countries for each year in the dataset?
15. Which five Asian countries had the highest life expectancy in 2007?
16. Calculate the total number of observations for each country in Europe. Help: use `n()` function.
17. How many observations do we have per continent?
18. Compute the average life expectancy by continent.
19. Rank countries according to their life expectancy and store it in a new column called `rank`. Rearrange the rows according to the ascending order of ranks (1, 2, 3...).
20. Calculate the mean and the standard error of the life expectancy for Belgium, Netherlands and France.
21. Categorize countries as “low” (`lifeExp < 50`) and “high” (`lifeExp > 50`) and store the values in a new column named “category”.