

# R Take-Home Assignment 4: World Development Data (Multiple Indicators)

Your Name

2025-10-01

## Overview

In this assignment you will explore a dataset combining three important development indicators for almost all countries in the world between 1990 and 2020:

- **GDP per capita** – a measure of economic output per person
- **Life expectancy** – an indicator of population health
- **CO<sub>2</sub> emissions per capita** – a measure of environmental impact

The dataset `world_data.csv` has already been prepared for you. It allows you to study how wealth, health, and environmental outcomes are related across countries and over time.

## Source

File on disk `world_data.csv` (should be stored in the same folder as this `.Rmd`)

## Skills

reading local CSVs, subsetting, cleaning, plotting, interpreting relationships

## Preparation

You will need: `dplyr`, `ggplot2`, `readr`.

```
## Task: load packages (hint: library())
## library(dplyr)
## library(ggplot2)
## library(readr)
```

## 1. Load the dataset

Before we can do any analysis, the dataset needs to be loaded into R. Since this file is stored locally, you'll also practise one of the most common first steps in data analysis: reading data from disk.

**Task:** Load `world_data.csv` into R and store it as `world_data`. Print the number of rows and columns, and look at the first few lines of the data.

```
## Task: read CSV (hint: read_csv("world_data.csv"))
## Task: check dimensions (hint: dim())
## Task: preview (hint: head())
```

## 2. Subset the dataset

Instead of analysing every year, it's common to take a snapshot of one year for cross-country comparisons. Here we'll use 2015, as it provides recent data but avoids issues with missing values in the latest years.

**Task:** Subset the dataset to the year 2015. Show the resulting table (a few rows is enough).

```
## Task: filter year == 2015 (hint: filter(year == 2015))
```

## 3. Explore summary statistics

Before making plots, it's useful to see some descriptive statistics. Averages and ranges can reveal whether values look reasonable, and whether there is a lot of variation between countries.

**Task:** For all countries in 2015, calculate the mean and range for GDP per capita, life expectancy, and CO<sub>2</sub> emissions per capita.

```
## Task: summarise mean and range for all 3 variables (hint: summarise(mean(...), range(...)))
```

## 4. Plot GDP vs life expectancy

A classic question in development studies is whether economic prosperity translates into better health outcomes. Plotting GDP per capita against life expectancy lets us explore this visually. Adding CO<sub>2</sub> emissions as colour allows us to consider whether higher prosperity comes with environmental costs.

**Task:** Create a scatterplot of GDP per capita (x-axis, log scale) vs life expectancy (y-axis). Colour the points by CO<sub>2</sub> emissions per capita. Add informative axis labels and a title.

```
## Task: plot scatter (hint: ggplot(..., aes(x=gdp_per_capita, y=life_expectancy, color=co2_per_capita))
```

## 5. Reflection

The final step in any analysis is to connect numbers and plots back to real-world meaning. Does wealth always mean health? Are there exceptions? Do you notice patterns in CO<sub>2</sub> emissions?

**Task:** Write 3–4 sentences interpreting your findings. Mention at least one interesting or surprising pattern.

*Write your reflection here...*