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_2 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.

- **File:** `world_data.csv` (stored in the same folder as this `.Rmd`)
- **Skills:** reading local CSVs, subsetting, cleaning, plotting, interpreting relationships
- **Deliverable:** knitted **HTML** plus your completed `.Rmd`

Preparation

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

```
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.

```
world_data <- read_csv("world_data.csv", show_col_types = FALSE)
print(dim(world_data))
```

```
## [1] 4322    5
```

```
head(world_data)
```

```
## # A tibble: 6 x 5
```

```
##   country      year gdp_per_capita co2_per_capita life_expectancy
##   <chr>       <dbl>          <dbl>          <dbl>           <dbl>
## 1 Afghanistan  1990          1085.          0.168            51.6
## 2 Afghanistan  1991           984.          0.156            51.3
## 3 Afghanistan  1992           955.          0.112            51.4
## 4 Afghanistan  1993           658.          0.1              51.3
## 5 Afghanistan  1994           487.          0.089            50.7
## 6 Afghanistan  1995           721.          0.083            51.0
```

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).

```
subset_data <- world_data %>% filter(year == 2015)
head(subset_data)
```

```
## # A tibble: 6 x 5
##   country      year gdp_per_capita co2_per_capita life_expectancy
##   <chr>       <dbl>          <dbl>          <dbl>           <dbl>
## 1 Afghanistan  2015          1856.          0.286            57.8
## 2 Albania      2015         10085.          1.63             77.4
## 3 Algeria      2015         13781.          4                77.3
## 4 Angola       2015          6014.          0.977            64.0
## 5 Argentina    2015         19559.          4.41             76.6
## 6 Armenia      2015         10087.          1.87             75.6
```

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_2 emissions per capita.

```
subset_data %>% summarise(
  mean_gdp = mean(gdp_per_capita), range_gdp = paste(range(gdp_per_capita), collapse='-'),
  mean_life = mean(life_expectancy), range_life = paste(range(life_expectancy), collapse='-'),
  mean_co2 = mean(co2_per_capita), range_co2 = paste(range(co2_per_capita), collapse='-')
)
```

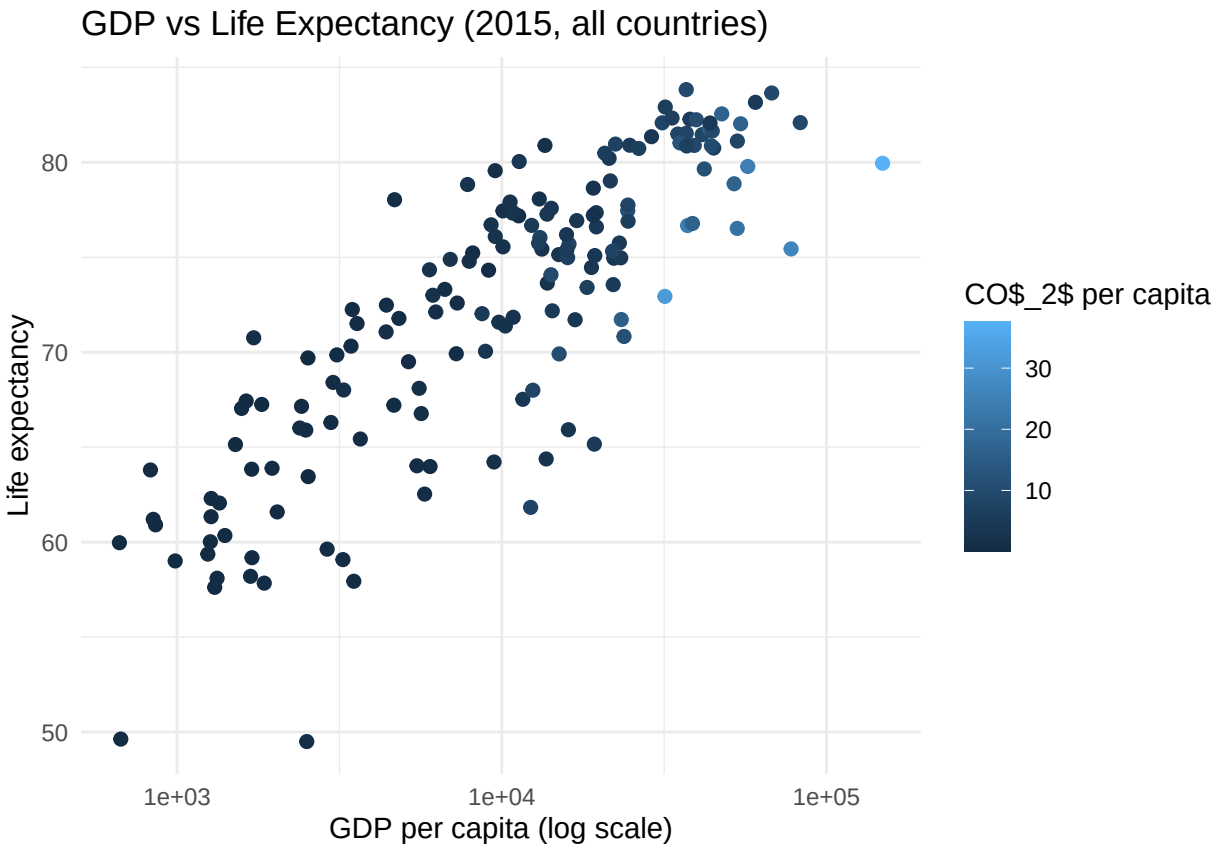
```
## # A tibble: 1 x 6
##   mean_gdp range_gdp                      mean_life range_life mean_co2 range_co2
##      <dbl> <chr>                              <dbl> <chr>         <dbl> <chr>
## 1   17638. 664.438548170731-148822.6354~      72.5 49.5-83.8~     4.77 0.032-37~
```

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_2 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_2 emissions per capita. Add informative axis labels and a title.

```r
ggplot(subset_data, aes(x=gdp_per_capita, y=life_expectancy, color=co2_per_capita)) +
  geom_point(size=2) +
  scale_x_log10() +
  labs(x='GDP per capita (log scale)', y='Life expectancy', color='CO$_2$ per capita',
       title='GDP vs Life Expectancy (2015, all countries)') +
  theme_minimal()
```



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_2 emissions?

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

Write your reflection here...