QTM350 Homework 5 - WDI Analysis

Yunshi Zhou

2026-02-24

Loading the Data

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.3     v readr     2.1.4
v forcats   1.0.0     v stringr   1.5.0
v ggplot2   3.4.3     v tibble    3.2.1
v lubridate 1.9.3     v tidyr     1.3.0
v purrr     1.0.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becor
```

```
wdi <- read_csv("wdi.csv")
```

```
Rows: 217 Columns: 14
-- Column specification --------------------------------------------------------
Delimiter: ","
chr  (1): country
dbl (13): inflation_rate, exports_gdp_share, gdp_growth_rate, gdp_per_capita...

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
glimpse(wdi)
```

```
Rows: 217
Columns: 14
$ country                          <chr> "Afghanistan", "Albania", "Algeria", "~
$ inflation_rate                   <dbl> NA, 6.725203, 9.265516, NA, NA, 21.355~
$ exports_gdp_share                <dbl> 18.38004, 37.39542, 31.44686, 46.95752~
$ gdp_growth_rate                  <dbl> -6.240172, 4.856402, 3.600000, 1.73501~
$ gdp_per_capita                   <dbl> 352.6037, 6810.1140, 5023.2529, 19673.~
$ adult_literacy_rate              <dbl> NA, 98.50000, NA, NA, NA, 72.40000, NA~
$ primary_school_enrolment_rate    <dbl> NA, 95.60671, 108.34393, NA, 90.14735,~
$ education_expenditure_gdp_share <dbl> NA, 2.749310, NA, NA, 2.666230, 2.3320~
$ measles_immunisation_rate        <dbl> 68, 86, 79, NA, 98, 37, 99, 83, 95, NA~
$ health_expenditure_gdp_share     <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
$ income_inequality                <dbl> NA, NA, NA, NA, NA, NA, NA, 40.7, 27.9~
$ unemployment_rate                <dbl> 14.100, 11.588, 12.437, NA, NA, 14.693~
$ life_expectancy                  <dbl> 62.87900, 76.83300, 77.12900, NA, NA, ~
$ total_population                 <dbl> 41128771, 2777689, 44903225, 44273, 79~
```

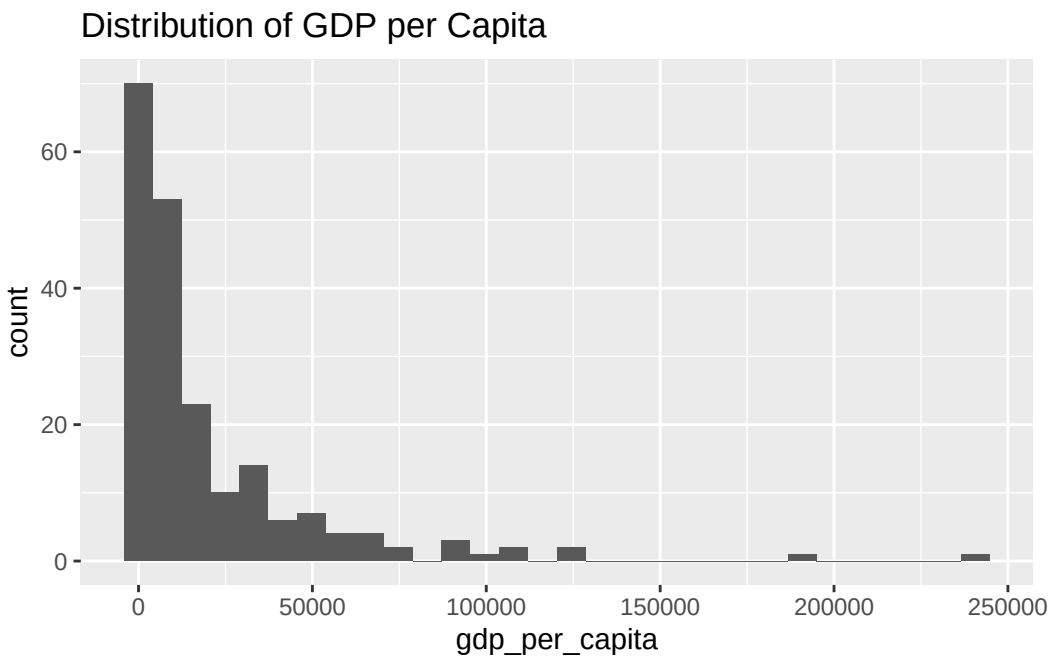
GDP Per Capita, Life Expectancy, and CO2 Emissions

```r
summary(wdi$gdp_per_capita)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
    259    2571    7588   20346   25983  240862      14
```

```r
ggplot(wdi, aes(x = gdp_per_capita)) +
  geom_histogram(bins = 30) +
  labs(title = "Distribution of GDP per Capita")
```

```
Warning: Removed 14 rows containing non-finite values (`stat_bin()`).
```



GDP per capita shows a highly right-skewed distribution. Most countries have relatively low income levels, while a small number of countries have extremely high values.

```
summary(wdi$life_expectancy)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  53.00   66.78   73.51   72.42   78.47   85.38       8
```

```
ggplot(wdi, aes(x = life_expectancy)) +
  geom_histogram(bins = 30) +
  labs(title = "Distribution of Life Expectancy")
```

```
Warning: Removed 8 rows containing non-finite values (`stat_bin()`).
```

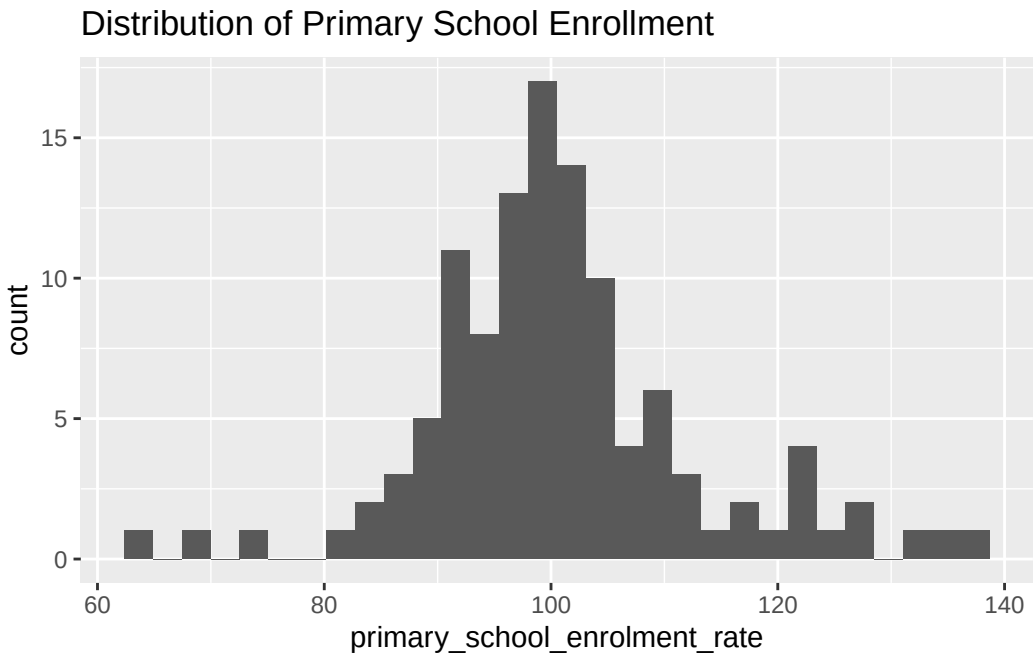
Life expectancy is more normally distributed. Most countries fall between 65 and 80 years, with the highest count as ~75 years old.

```
summary(wdi$primary_school_enrolment_rate)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  64.40   94.19  100.02  100.87  105.04  138.19     103
```

```
ggplot(wdi, aes(x = primary_school_enrolment_rate)) +
  geom_histogram(bins = 30) +
  labs(title = "Distribution of Primary School Enrollment")
```

```
Warning: Removed 103 rows containing non-finite values (`stat_bin()`).
```



Primary school enrollment rates are generally high across countries. Most countries show enrollment rates around 80-110%. (I don't know why the enrollment rate can be higher than 100%)

Relationship between Income and Health, Top Countries by Primary School Enrollment

```
Warning: Removed 20 rows containing missing values (`geom_point()`).
```

Summary: Statistics Table

Table 1: Table 1: Key Summary Statistics of Selected Indicators (Source: World Development Indicators)

GDP per Capita Mean	GDP per Capita Median	Life Expectancy Mean	Life Expectancy Median	Enrollment Rate Mean	Enrollment Rate Median
20345.71	7587.59	72.42	73.51	100.87	100.02

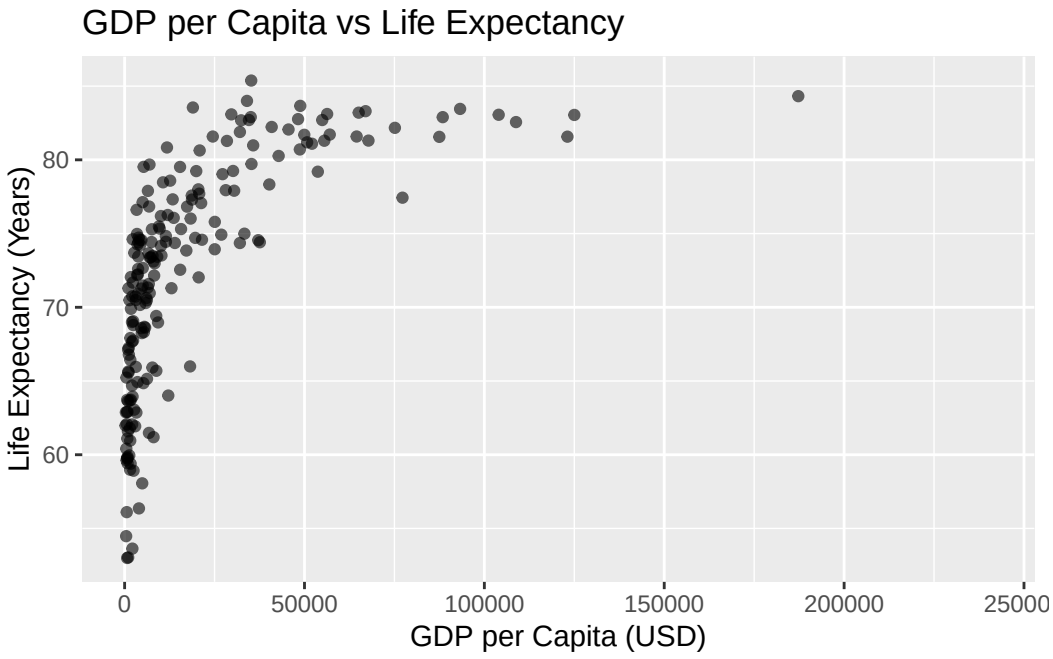


Figure 1: Figure 1: Relationship between GDP per capita and Life Expectancy (Source: World Development Indicators).

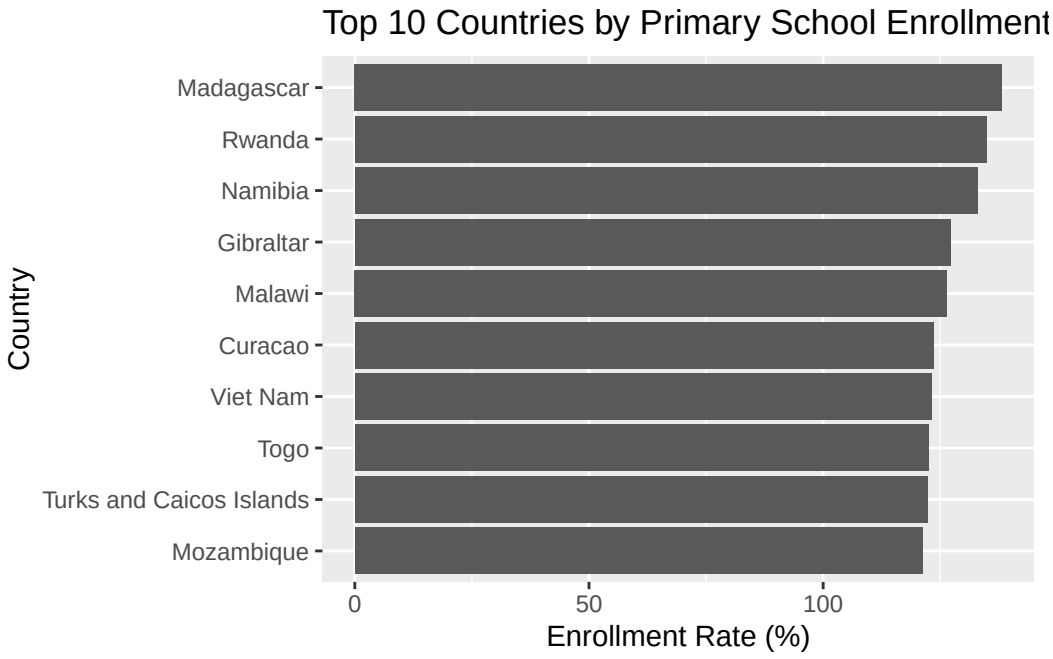


Figure 2: Figure 2: Top 10 Countries by Primary School Enrollment Rate (Source: World Development Indicators).