World-Development-Project

Max Jiang

2024-10-07

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   3.5.1     v tibble    3.2.1
v lubridate 1.9.3     v tidyr     1.3.1
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 beco
```

```
wdi <- read_csv('/Users/maxjiang/Desktop/QTM350/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.
```

```
head(wdi)
```

```
# A tibble: 6 x 14
  country        inflation_rate exports_gdp_share gdp_growth_rate gdp_per_capita
```

```
  <chr>                  <dbl>             <dbl>          <dbl>          <dbl>
1 Afghanistan            NA                 18.4          -6.24           353.
2 Albania                 6.73              37.4           4.86          6810.
3 Algeria                 9.27              31.4           3.60          5023.
4 American Samoa         NA                 47.0           1.74         19673.
5 Andorra                NA                 NA             9.56         42351.
6 Angola                 21.4               44.4           3.05          2933.
# i 9 more variables: adult_literacy_rate <dbl>,
#   primary_school_enrolment_rate <dbl>, education_expenditure_gdp_share <dbl>,
#   measles_immunisation_rate <dbl>, health_expenditure_gdp_share <dbl>,
#   income_inequality <dbl>, unemployment_rate <dbl>, life_expectancy <dbl>,
#   total_population <dbl>
```

```r
# Select three indicators for analysis
indicators <- wdi %>%
  select(gdp_per_capita = `gdp_per_capita`,
         life_expectancy = `life_expectancy`,
         total_population = `total_population`)
summary(indicators)
```

```
 gdp_per_capita   life_expectancy total_population
 Min.   :   259   Min.   :53.00   Min.   :1.131e+04
 1st Qu.:  2571   1st Qu.:66.78   1st Qu.:8.087e+05
 Median :  7588   Median :73.51   Median :6.465e+06
 Mean   : 20346   Mean   :72.42   Mean   :3.654e+07
 3rd Qu.: 25983   3rd Qu.:78.47   3rd Qu.:2.607e+07
 Max.   :240862   Max.   :85.38   Max.   :1.417e+09
 NA's   :14       NA's   :8
```

Exploratory Data Analysis: We are analyzing three key indicators: GDP per capita, life expectancy, and total population. Below are the summary statistics for these indicators:

```r
summary(indicators)
```


```
 gdp_per_capita   life_expectancy total_population
 Min.   :   259   Min.   :53.00   Min.   :1.131e+04
 1st Qu.:  2571   1st Qu.:66.78   1st Qu.:8.087e+05
 Median :  7588   Median :73.51   Median :6.465e+06
 Mean   : 20346   Mean   :72.42   Mean   :3.654e+07
 3rd Qu.: 25983   3rd Qu.:78.47   3rd Qu.:2.607e+07
 Max.   :240862   Max.   :85.38   Max.   :1.417e+09
 NA's   :14       NA's   :8
```

The summary statistics for GDP per capita, life expectancy, and total population reveal significant disparities among countries. GDP per capita shows a wide range, from 259 to 240,862, with a median of 7,588, indicating a right-skewed distribution where a few countries have exceptionally high GDP per capita. This is evident as the mean (20,346) is significantly higher than the median. Life expectancy also varies considerably, ranging from 53 to 85 years, with a median of 73.51, which is slightly higher than the mean (72.42), indicating a relatively symmetric distribution but slightly left-skewed. The total population displays the most extreme variation, ranging from just over 11,000 to 1.417 billion, with a median of 6.46 million. This distribution is heavily skewed due to a few highly populous countries, as the mean population (36.54 million) is much higher than the median. Additionally, missing values for GDP per capita (14) and life expectancy (8) indicate incomplete data for some countries, which should be addressed in further analysis. Overall, the statistics reflect significant socioeconomic differences across the dataset, with notable outliers in wealth and population.

```
# Density plot for GDP per capita
ggplot(indicators, aes(x = gdp_per_capita)) +
  geom_density(fill = "blue", alpha = 0.7) +
  labs(title = "Density Plot of GDP Per Capita", x = "GDP Per Capita", y = "Density") +
  theme_minimal()
```

```
Warning: Removed 14 rows containing non-finite outside the scale range
(`stat_density()`).
```

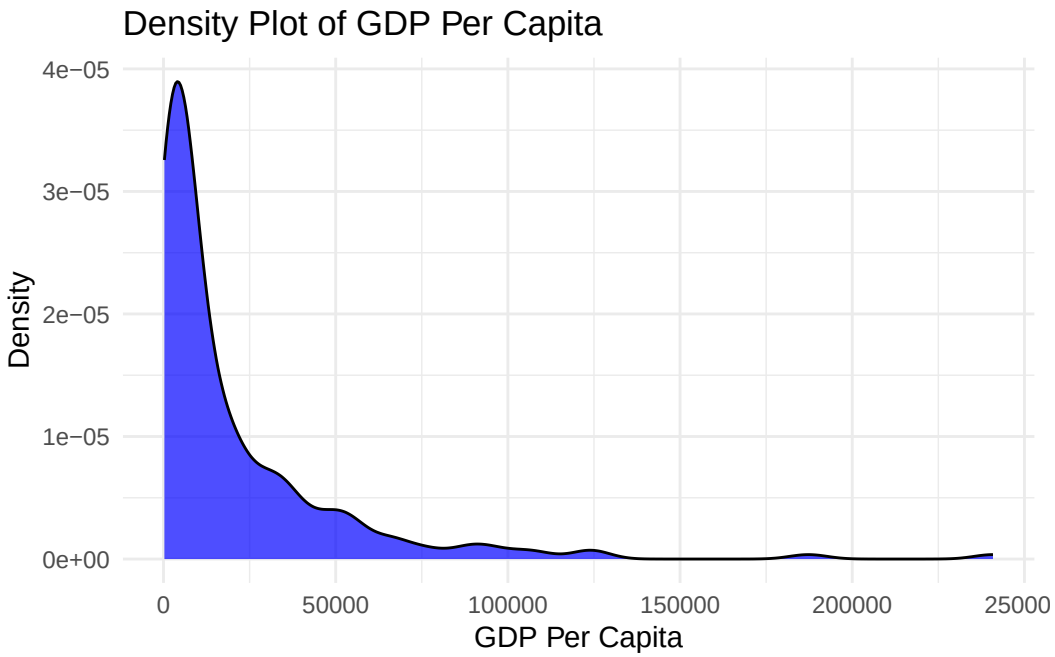

Figure 1: Density Plot of GDP Per Capita from the World Development Indicators dataset

```
# Scatter plot for life expectancy vs GDP per capita
ggplot(indicators, aes(x = gdp_per_capita, y = life_expectancy)) +
  geom_point(color='purple') +
  labs(title = "Life Expectancy vs GDP Per Capita", x = "GDP Per Capita", y = "Life Expectanc
  theme_minimal()
```

```
Warning: Removed 20 rows containing missing values or values outside the scale range
(`geom_point()`).
```

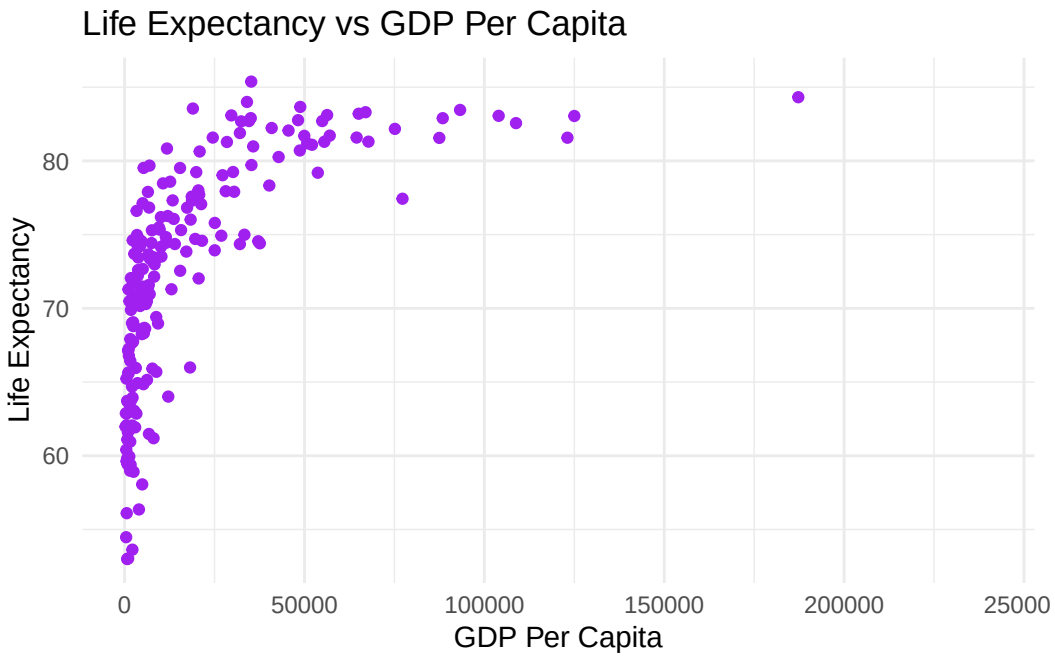

Figure 2: Scatter Plot of Life Expectancy vs GDP Per Capita from the World Development Indicators dataset

Summary Table of Key Indicators

```
knitr::kable(summary(indicators), caption = "Summary of Key Indicators")
```

Table 1: Summary of Key Indicators

gdp_per_capita	life_expectancy	total_population
Min. : 259	Min. :53.00	Min. :1.131e+04
1st Qu.: 2571	1st Qu.:66.78	1st Qu.:8.087e+05
Median : 7588	Median :73.51	Median :6.465e+06
Mean : 20346	Mean :72.42	Mean :3.654e+07
3rd Qu.: 25983	3rd Qu.:78.47	3rd Qu.:2.607e+07
Max. :240862	Max. :85.38	Max. :1.417e+09
NA's :14	NA's :8	NA

Cross-References

As shown in Figure 1, the distribution of GDP per capita across countries is highly skewed. Additionally, Figure 2 demonstrates the relationship between life expectancy and GDP per capita, indicating a possible positive correlation between the two.

The key statistics for the selected indicators, including GDP per capita, life expectancy, and total population, are summarized in Table 1.

The data used in this analysis comes from the World Bank Development Indicators dataset (Bank 2024). Additionally, the analysis techniques applied here are informed by concepts discussed in *Data Science for Business* by Provost and Fawcett (Provost and Fawcett 2013).

Bank, World. 2024. "World Bank Development Indicators." https://data.worldbank.org/indicator.

Provost, Foster, and Tom Fawcett. 2013. *Data Science for Business*. O'Reilly Media.