Data Analysis Presentation

Ada

2025-02-22

Data Analysis Presentation

Tasks

1. Please initialise a new `.qmd` file with an appropriate `YAML` header. Include metadata such as `title`, `author`, `date`, and specify the output format as `HTML` and `PDF`.
2. Load the dataset using your preferred programming language (R or Python).

```
-- 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 becor

Attaching package: 'kableExtra'


The following object is masked from 'package:dplyr':

    group_rows
```

3. Conduct exploratory data analysis on at least three indicators of your choice. Summarise your findings in markdown sections. Show your code and results.

```
'data.frame':   217 obs. of  14 variables:
 $ country                          : chr  "Afghanistan" "Albania" "Algeria" "American Samoa"
 $ inflation_rate                   : num  NA 6.73 9.27 NA NA ...
 $ exports_gdp_share                : num  18.4 37.4 31.4 47 NA ...
 $ gdp_growth_rate                  : num  -6.24 4.86 3.6 1.74 9.56 ...
 $ gdp_per_capita                   : num  353 6810 5023 19673 42351 ...
 $ adult_literacy_rate              : num  NA 98.5 NA NA NA ...
 $ primary_school_enrolment_rate    : num  NA 95.6 108.3 NA 90.1 ...
 $ education_expenditure_gdp_share: num  NA 2.75 NA NA 2.67 ...
 $ measles_immunisation_rate        : num  68 86 79 NA 98 37 99 83 95 NA ...
 $ health_expenditure_gdp_share     : num  NA NA NA NA NA NA NA NA NA NA ...
 $ income_inequality                : num  NA NA NA NA NA NA NA 40.7 27.9 NA ...
 $ unemployment_rate                : num  14.1 11.6 12.4 NA NA ...
 $ life_expectancy                  : num  62.9 76.8 77.1 NA NA ...
 $ total_population                 : num  41128771 2777689 44903225 44273 79824 ...
```

Inflation Rate Analysis

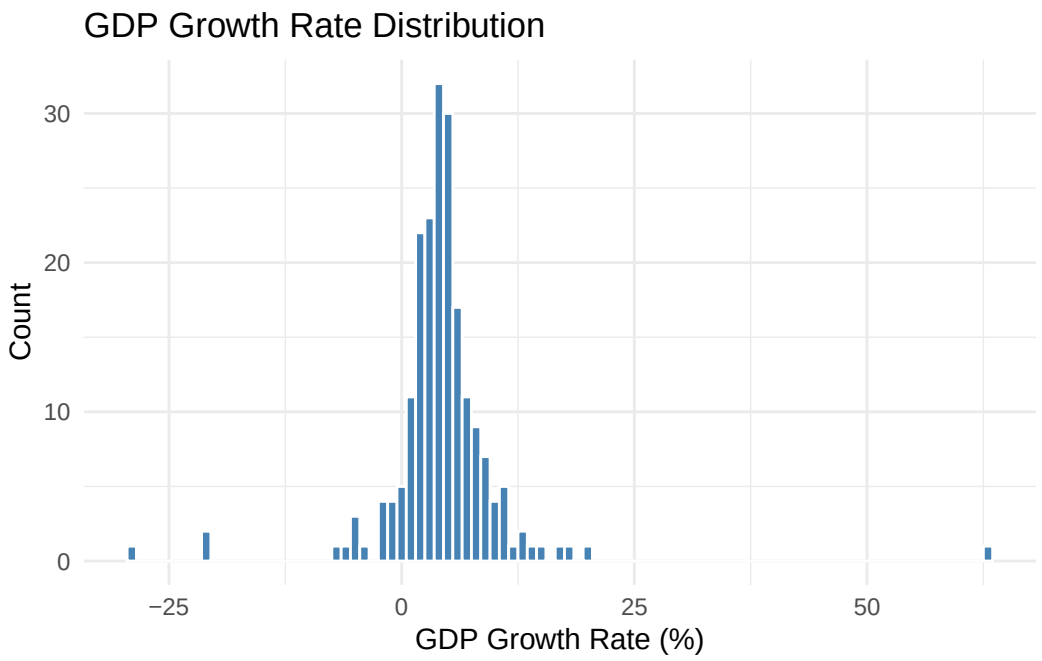
Mean	Median	SD
12.49394	7.967574	19.68243

The inflation rate exhibits a right-skewed distribution, with a mean of approximately 12.44, a median of about 7.97, and a standard deviation of around 19.68, indicating that the inflation rate distribution in this sample is relatively dispersed and shows significant variability.

GDP Growth Rate Distribution

Mean	Median	SD
4.368901	4.204431	6.62681

```
Warning: Removed 15 rows containing non-finite outside the scale range
(`stat_bin()`).
```



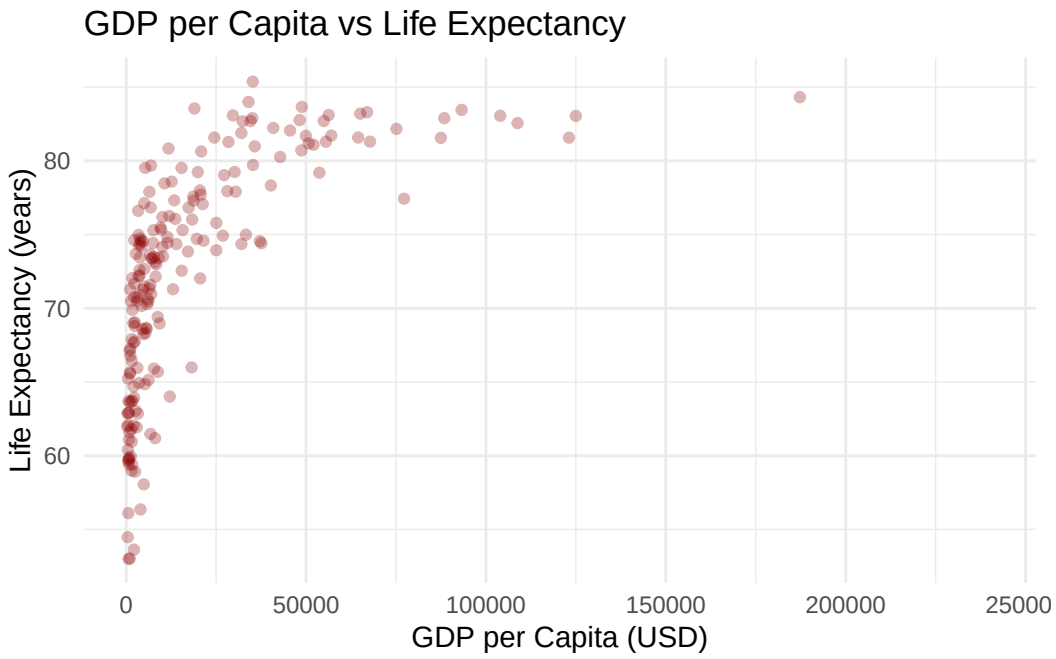
The GDP growth rate appears to be nearly symmetrical or slightly right-skewed, given that the mean (approximately 4.37) is slightly higher than the median (about 4.20). The standard deviation of around 6.63 suggests moderate variability in the GDP growth rates across the sample.

Relationship between GDP per Capita and Life Expectancy

```
Call:
lm(formula = life_expectancy ~ log(gdp_per_capita), data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-13.141  -2.163   0.497   2.545   8.761

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)          32.525      1.853   17.55   <2e-16 ***
log(gdp_per_capita)   4.460      0.205   21.75   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.184 on 195 degrees of freedom
```

```
  (20 observations deleted due to missingness)
Multiple R-squared:  0.7081,    Adjusted R-squared:  0.7066
F-statistic: 473.1 on 1 and 195 DF,  p-value: < 2.2e-16
```

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



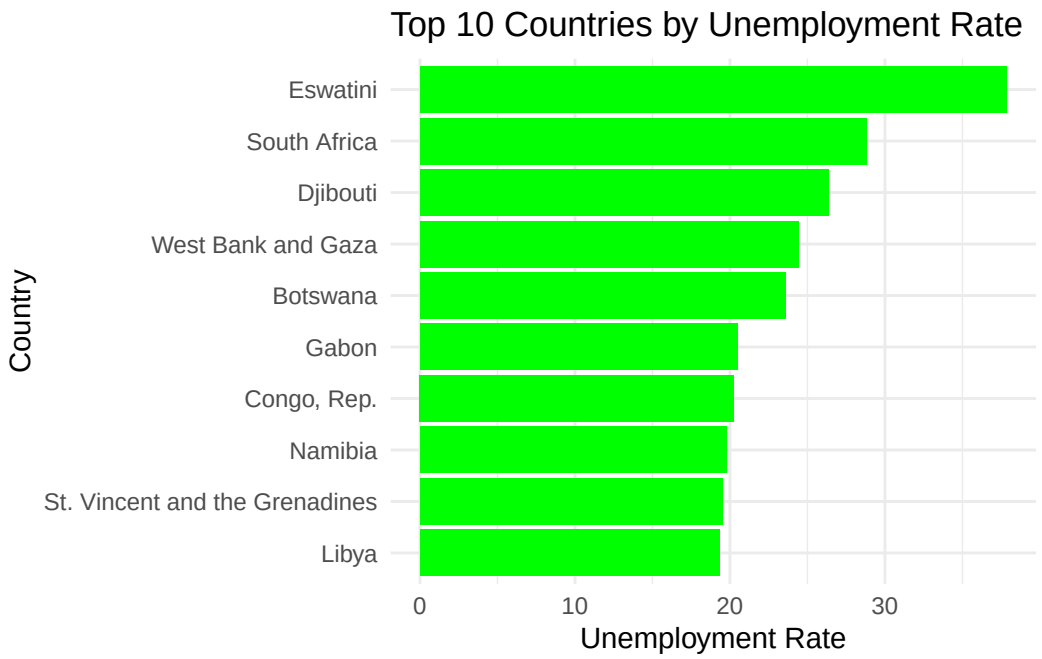
Higher GDP per capita is strongly associated with longer life expectancy. The log-linear model indicates that a one-unit increase in the natural log of GDP per capita corresponds to an increase of about 4.46 years in life expectancy, demonstrating diminishing returns as income rises.

Summary Table

Mean_GDP	Median_GDP	Mean_LifeExp	Median_LifeExp
20345.71	7587.588	72.41652	73.51463

4. Create at least two different types of plots (e.g., bar chart, scatter plot) to represent your analysis. Use Quarto code chunks to embed these visualisations. Add a title and axis labels to each plot. Use Quarto to include a caption and a reference to the source of the data. Hide your code in the final document.

Other plots




```
Warning: Removed 15 rows containing non-finite outside the scale range
(`stat_ydensity()`).
```

```
Warning: Removed 15 rows containing non-finite outside the scale range
(`stat_boxplot()`).
```

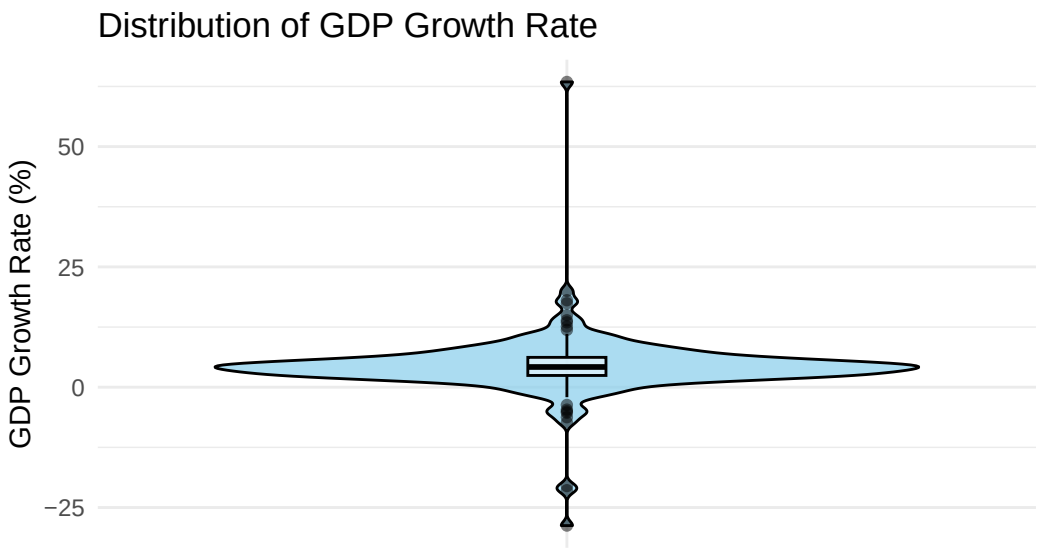


Figure @fig:violin−plot: Violin plot showing the distribution of GDP growth rate.

Key statistics

6. Construct a table that highlights some key statistics from your analysis. Ensure the table is well-formatted and included in the report.

```
Warning: Since gt v0.3.0, `columns = vars(...)` has been deprecated.
* Please use `columns = c(...)` instead.
```

Key Statistics from the Dataset

Summary of GDP per Capita, Life Expectancy, and Unemployment Rat

Average GDP per Capita	Median GDP per Capita	Average Life Expectancy	Median Life Expecta
20,345.71	7,587.59	72.42	7

6. Include cross-references to your figures and tables within the text. Demonstrate proper labeling and referencing techniques.

Table 5: Summary statistics for GDP growth rate

Mean	Median	SD
4.368901	4.204431	6.62681

The key summary statistics can be found in Table Table 5.

References

Add a bibliography using BibTeX (.bib). Cite at least two sources related to your analysis.

Our analysis uses data obtained from the World Development Indicators provided by the World Bank (World Bank 2025a). Using methods from (World Bank 2025b) provided by the World Bank.

Summary

The inflation rate exhibits a right-skewed distribution, with a mean of approximately 12.44, a median of about 7.97, and a standard deviation of around 19.68, indicating that the inflation rate distribution in this sample is relatively dispersed and shows significant variability.

The GDP growth rate appears to be nearly symmetrical or slightly right-skewed, given that the mean (approximately 4.37) is slightly higher than the median (about 4.20). The standard deviation of around 6.63 suggests moderate variability in the GDP growth rates across the sample.

Higher GDP per capita is strongly associated with longer life expectancy. The log-linear model indicates that a one-unit increase in the natural log of GDP per capita corresponds to an increase of about 4.46 years in life expectancy, demonstrating diminishing returns as income rises.

World Bank. 2025a. “World Development Indicators.”
———. 2025b. “World Development Indicators: Methodology.”