Variable: Country Name, Country Code

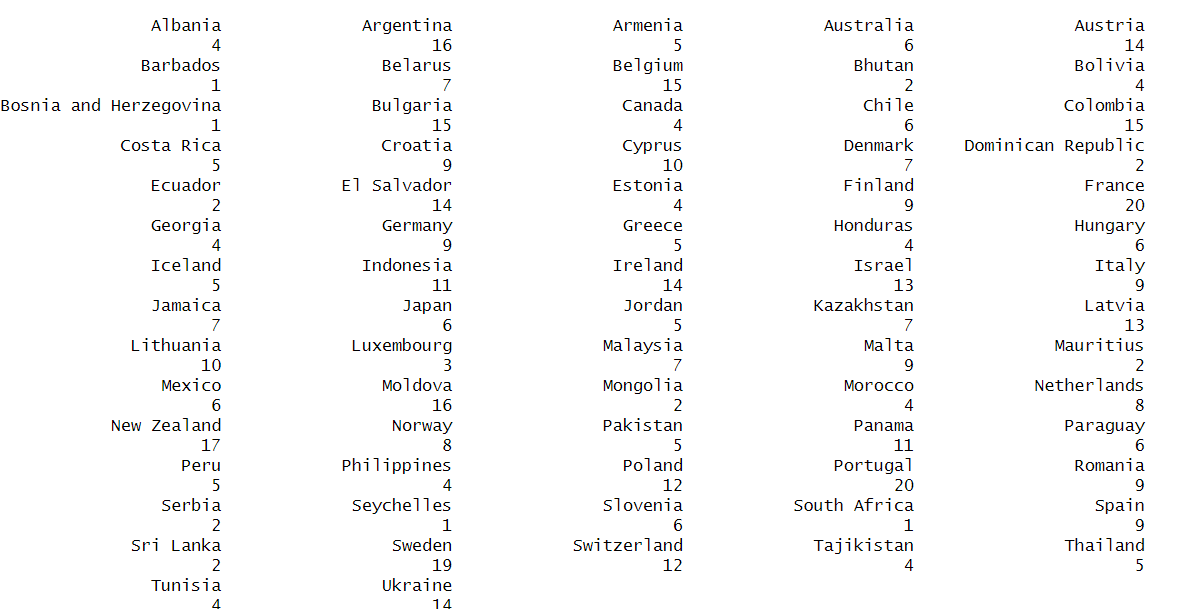
Original data file: All data files

Original variable name: Country Name, Country Code

Number of Observations: 558

No missing value

Definition: “Country Name” indicates the name of the country, and the “Country Code” variable is a unique 3-letter code that indicates the country. Because these two are the same variables, we will use the same frequency table (no point in using a histogram because of the high number of countries):



From the frequency table above, we can see that data completeness varies significantly by country. France has the highest number of observations (20), whereas several countries have only one observation.

Variable: year

Original data file: All data files

Original variable name: year (or Year)

Number of Observations: 558

No missing value

Definition: The year of the observation

Mean: 2,005.11

Standard Deviation: 7.145

Minimum: 1,980

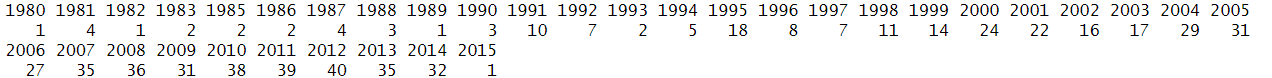
Median: 2,007

Maximum: 2,015

25th percentile: 2,001

75th percentile: 2,011

Frequency table and boxplot(because boxplot makes more sense than histogram in terms of showing the distribution)





The boxplot above shows that the distribution of observation years is highly skewed toward more recent times. This is because datasets from more recent years are more complete, making observations from later periods more likely to be retained when taking the intersection of multiple datasets. This may reduce the representativeness of the results for earlier periods.

Variable: region\_wb

Original data file: wiidcountry.xlsx

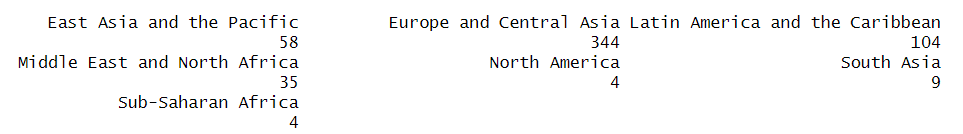
Original variable name: region\_wb

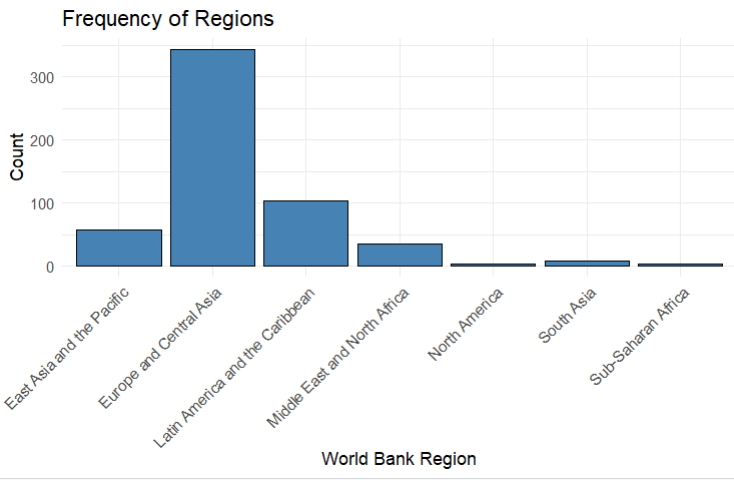
Number of Observations: 558

No missing value

Definition: The region of the country defined by the World Bank

Frequency Table and Histogram:





There are significantly more observations for the Europe and Central Asia region compared to other regions, while North America, South Asia, and Sub-Saharan Africa have very few observations. This imbalance makes it difficult to run grouped regressions for regions with limited data. However, the large number of observations for Europe and Central Asia allows for region-specific regressions, and the results for this group should be robust.

Variable: region\_un\_sub

Original data file: wiidcountry.xlsx

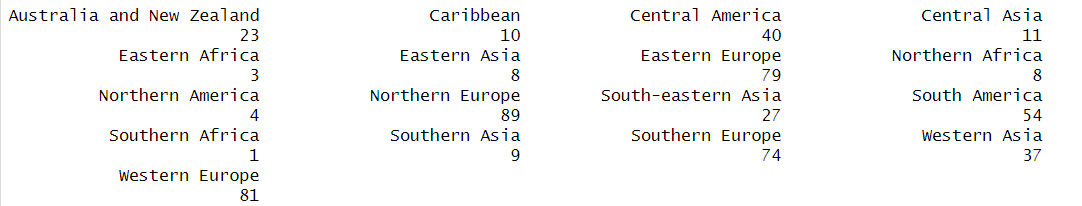
Original variable name: region\_un\_sub

Number of Observations: 558

No missing value

Definition: The subregion of the country defined by the United Nations

Frequency Table and Histogram:



This variable complements region\_wb by providing a more detailed regional classification. For the region\_wb variable, we observe that Europe and Central Asia have a relatively high number of observations. However, this variable reveals that observations from Central Asia are relatively limited, while Northern, Western, and Southern Europe have similar numbers of observations, all noticeably higher than those for Eastern Europe.

Variable: region\_un\_sub

Original data file: wiidcountry.xlsx

Original variable name: region\_un\_sub

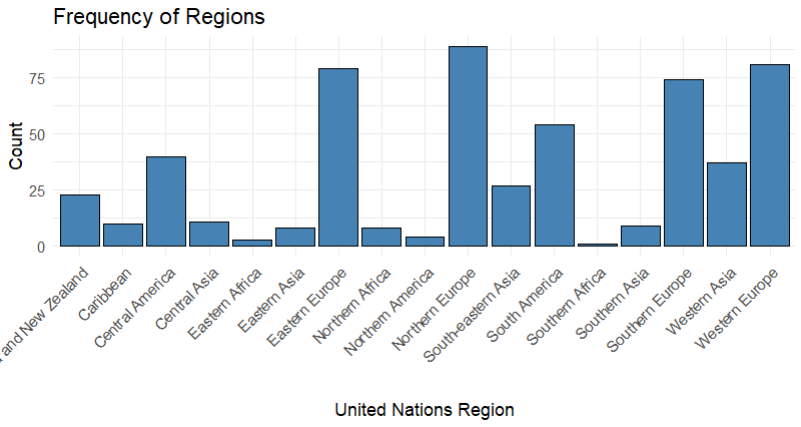
Number of Observations: 558

No missing value

Definition: The subregion of the country defined by the United Nations

Frequency Table and Histogram:





This variable complements region\_wb by providing a more detailed regional classification. For the region\_wb variable, we observe that Europe and Central Asia have a relatively high number of observations. However, this variable reveals that observations from Central Asia are relatively limited, while Northern, Western, and Southern Europe have similar numbers of observations, all noticeably higher than those for Eastern Europe. Additionally, the number of observations from South America is also relatively high, even after excluding Brazil due to data issues.

Variable: incomegroup

Original data file: wiidcountry.xlsx

Original variable name: incomegroup

Number of Observations: 558

No missing value

Definition: What incomegroup a country belongs to defined by the United Nations

Frequency Table and Histogram:





We can see that there are far more observations from high-income countries than from upper/lower-middle-income countries. Also, there are no Low-income observations, which are included as one of the categories. Again, this is another possible influence due to the incompleteness of the data. Because the unbalanced number of observations, this variable is not used as a control variable.

Variable: Remittance\_as\_percent

Original data file: personal\_remittance\_received\_percentage.csv

Original variable name: Personal remittances, received (% of GDP) (transformed to panel data)

Number of Observations: 558

No missing value

Definition: Personal remittances, received in % of GDP

Mean: 3.335

Standard Deviation: 5.933

Minimum: 0.000

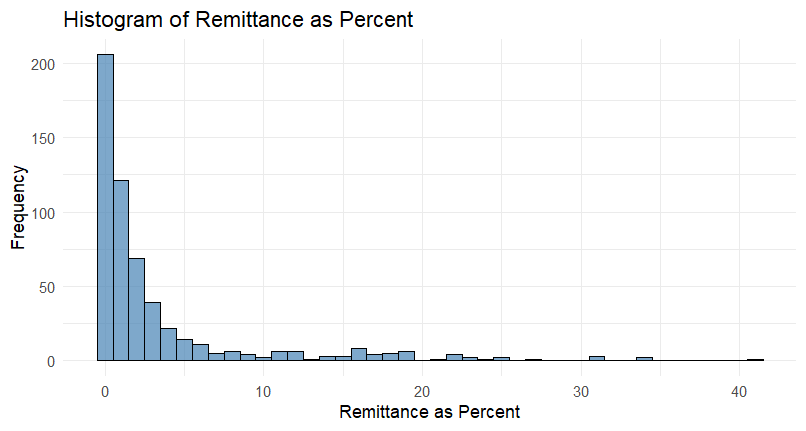
Median: 0.952

Maximum: 40.696

25th percentile: 0.282

75th percentile: 2.906

Histogram:



The histogram and descriptive statistics show that remittances as a percentage of GDP are highly left-skewed, with nearly 75% of observations below 2%. However, some observations exceed 10%. This wide range should be taken into account when running regressions. We observe outliers of this variable, most of them are different years of Jordan.

Variable: Remittance\_recieved

Original data file: personal\_remittance\_received.csv

Original variable name: Personal remittances, received (current US$) (transformed to panel data)

Number of Observations: 558

No missing value

Definition: Personal remittances, received (current US$)

Mean: 2,816,277,880

Standard Deviation: 4,515,922,802

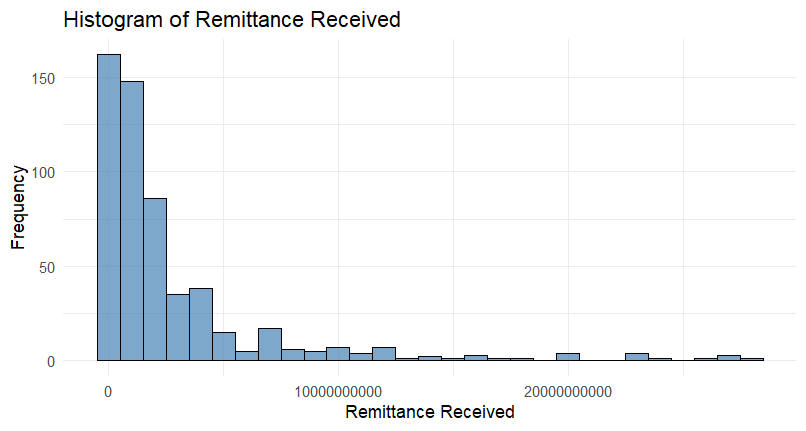
Minimum: 0.000

Median: 1,176,385,748

Maximum: 27,976,008,592

25th percentile: 427,102,041

75th percentile: 3,189,784,018

Histogram:

In general, the distribution of remittance values has a higher variance compared to remittance as a percentage of GDP. However, it still has a relatively low variance, since it still has a high Standard Deviation, especially compared to the mean. France and Mexico are the main outliers of this distribution.

Variable: Remittance\_paid

Original data file: personal\_remittance\_paid.csv

Original variable name: Personal remittances, paid (current US$) (transformed to panel data)

Number of Observations: 558

No missing value

Definition: Personal remittances, paid (current US$)

Mean: 1,814,013,631

Standard Deviation: 3,603,876,547

Minimum: 0.000

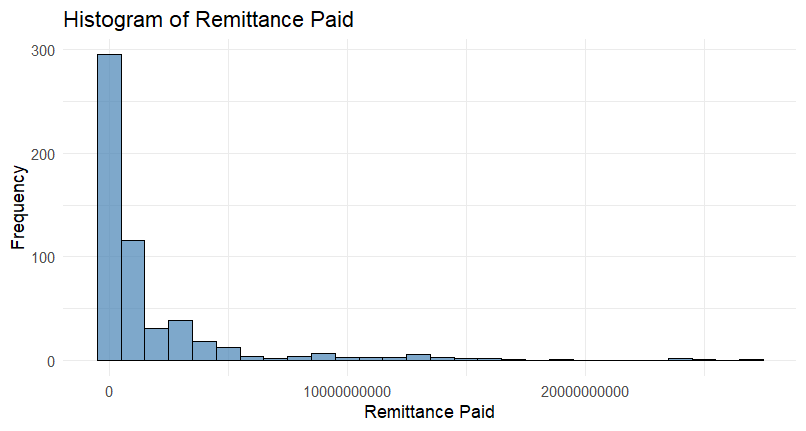
Median: 389,359,191

Maximum: 26,816,169,395

25th percentile: 95,014,228

75th percentile: 1,680,397,013

Histogram:



Similar to the two Remittance-related variables above, the value of Remittance received also is right skewed, has high variance, and outliers (mainly Switzerland).

Variable: gini\_std

Original data file: wiidcountry.xlsx

Original variable name: gini\_std

Number of Observations: 558

No missing value

Definition: Standardized Gini coefficient

Mean: 37.013

Standard Deviation: 8.693

Minimum: 22.190

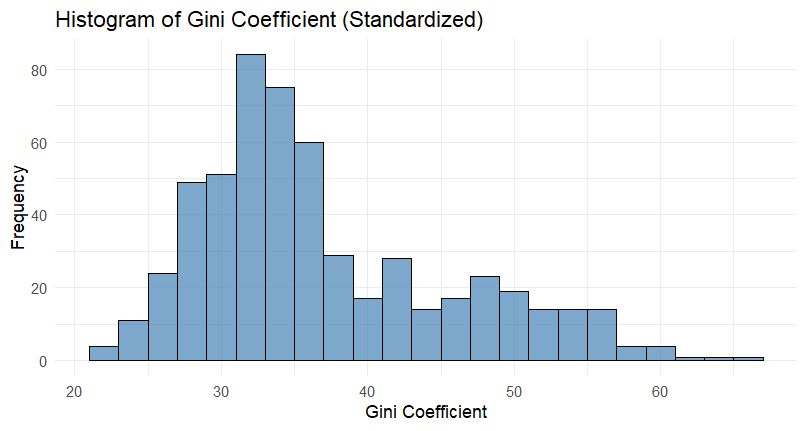
Median: 34.465

Maximum: 66.660

25th percentile: 31.060

75th percentile: 41.752

Histogram:



Compared to 3 Remittance variables, the gini coefficient is less skewed (although still right-skewed), with a smaller standard deviation (compared to its mean). It still has some outliers, consisting of countries like Bolivia and Peru (higher inequality).

Variable: bottom20

Original data file: wiidcountry.xlsx

Original variable name: bottom20

Number of Observations: 558

No missing value

Definition: Income share of the bottom 20%

Mean: 6.622

Standard Deviation: 2.031

Minimum: 1.008

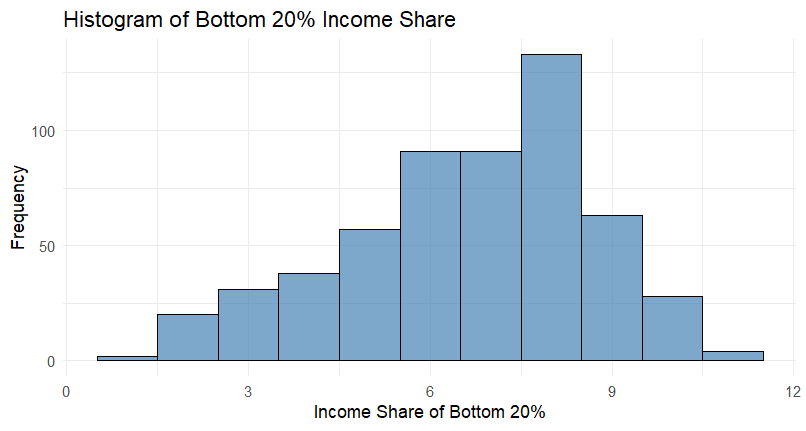
Median: 6.950

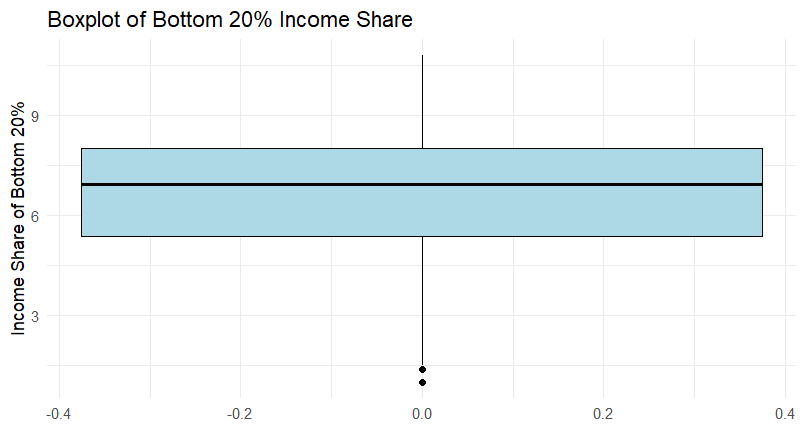
Maximum: 10.800

25th percentile: 5.389

75th percentile: 8.028

Histogram and box plot:





The income share of the bottom 20% is slightly left-skewed, indicating that most values are concentrated on the higher end, with a few countries having significantly lower income shares. There are two outliers close to 0, representing extreme income inequality (low share of income).

Variable: top20

Original data file: wiidcountry.xlsx

Original variable name: top20

Number of Observations: 558

No missing value

Definition: Income share of the top 20%

Mean: 37.029

Standard Deviation: 3.772

Minimum: 30.284

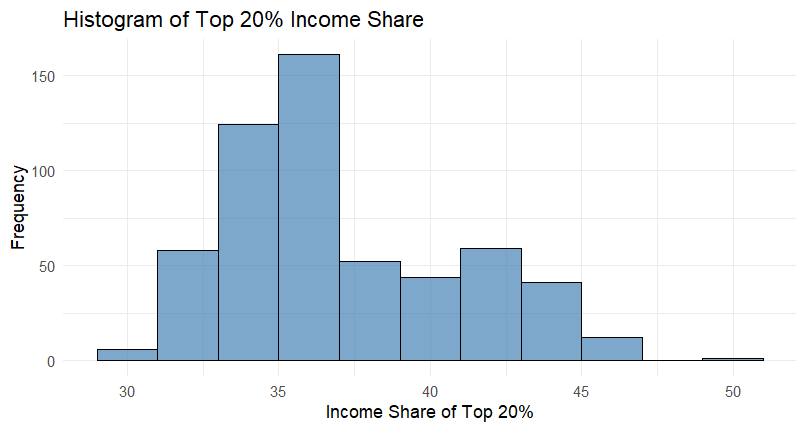
Median: 35.961

Maximum: 50.291

25th percentile: 34.494

75th percentile: 39.606

Histogram and boxplot:





The income share of the top 20% is slightly right-skewed. The variation of this variable is not high because its standard deviation is relatively small. Some outliers indicate a very high share of income by the top 20% earners, with the highest one being South Africa 2012.

Variable: school\_13

Original data file: school\_expectancy\_13.csv

Original variable name: Observation Value

Number of Observations: 558

No missing value

Definition: The school expectancy of ISCED levels 1 to 3 (equivalent to from primary school to high school).

Mean: 11.970

Standard Deviation: 1.434

Minimum: 6.254

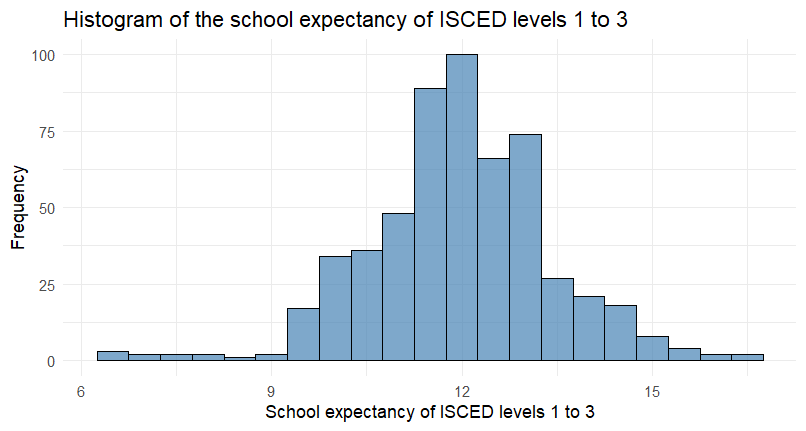
Median: 11.982

Maximum: 16.462

25th percentile: 11.205

75th percentile: 12.857

Histogram and boxplot:



The distribution of the school expectancy is close to a normal distribution as most observations have values around 12 and low variation. There are some outliers with low school expectancy, mainly from Pakistan and Morocco.

Variable: school\_58

Original data file: school\_expectancy\_58.csv

Original variable name: Observation Value

Number of Observations: 558

No missing value

Definition: The school expectancy of ISCED levels 5 to 8 (equivalent to from short-cycle tertiary education to Doctorate or equivalent).

Mean: 2.494

Standard Deviation: 1.012

Minimum: 0.158

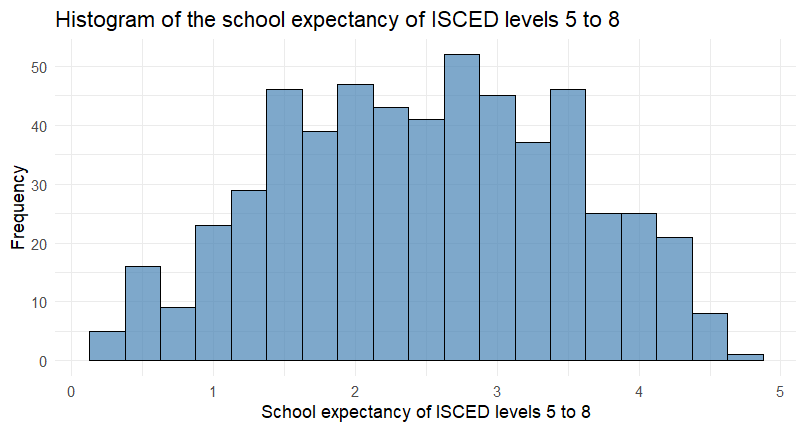
Median: 2.504

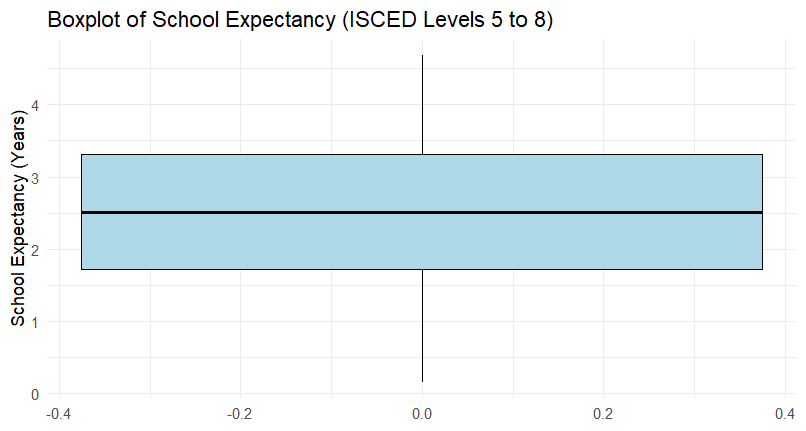
Maximum: 4.682

25th percentile: 1.719

75th percentile: 3.311

Histogram and boxplot:





The distribution of the school expectancy of ISCED levels 5 to 8 is approximately symmetric, with most values concentrated around 2 to 3 years. From the boxplot, we observe that there is no outlier.

Variable: Real GDP

Original data file: WEO\_Oct2024.xlsx

Original variable name: Gross domestic product, constant prices (Converted using the Gross domestic product, deflator. In panel data form)

Number of Observations: 558

No missing value

Definition: The real GDP expressed in terms of constant Purchasing power parity

Mean: 647.081

Standard Deviation: 1,117.17

Minimum: 0.104

Median: 293.397

Maximum: 14,250.20

25th percentile: 79.964

75th percentile: 609.840

Boxplot:



From the boxplot, we can see that real GDP is highly right-skewed and has a very high standard deviation. Although we convert nominal GDP to real GDP expressed in constant purchasing power parity, the GDP gap between different countries remains significant.

Variable: GDP.per.capita

Original data file: WEO\_Oct2024.xlsx

Original variable name: Gross domestic product per capita, current prices

Number of Observations: 558

No missing value

Definition: The GDP per capita expressed in GDP in PPP dollars per person

Mean: 31,322.77

Standard Deviation: 19,757.33

Minimum: 1,753.93

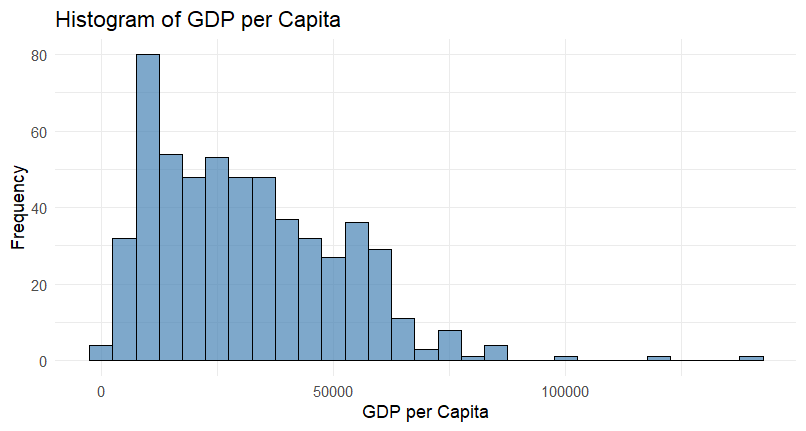
Median: 28,166.70

Maximum: 138,915.89

25th percentile: 14,728.51

75th percentile: 44,529.87

Histogram and Boxplot:





In general, the distribution of GDO per Capita is slightly right-skewed. The standard deviation is relatively low. However, these are three outliers, which are three observations of Luxembourg 2008.

Variable: Total investment

Original data file: WEO\_Oct2024.xlsx

Original variable name: Total investment

Number of Observations: 558

No missing value

Definition: expressed as a ratio of total investment in the current local currency and GDP in current local currency.

Mean: 23.834

Standard Deviation: 5.820

Minimum: 9.965

Median: 22.968

Maximum: 70.164

25th percentile: 20.564

75th percentile: 25.929

Boxplot:



The distribution of total investment(in ratio) is right-skewed. It has a wide range, but a relatively low standard deviation. Also, the mean and median are relatively close, which shows that most observations have value around these two numbers. There are outliers at both ends.

Variable: Gross national savings

Original data file: WEO\_Oct2024.xlsx

Original variable name: Gross national savings

Number of Observations: 558

No missing value

Definition: Expressed as a ratio of gross national savings in current local currency and GDP in current local currency.

Mean: 21.408

Standard Deviation: 7.297

Minimum: -57.542

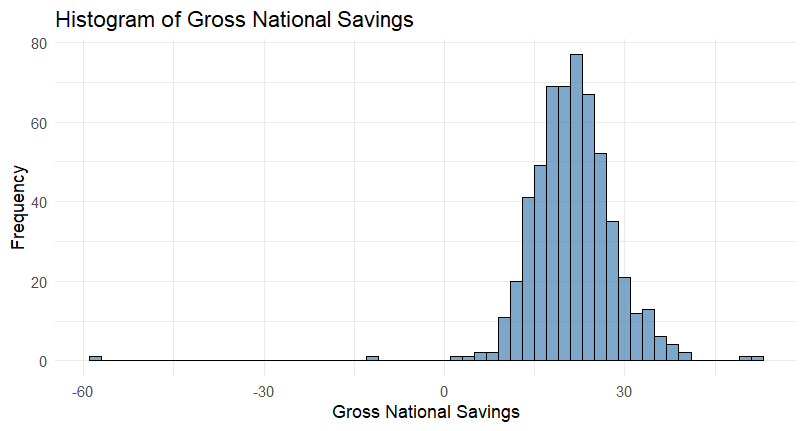
Median: 21.199

Maximum: 52.961

25th percentile: 17.359

75th percentile: 25.368

Histogram:



The distribution of Gross National Savings is approximately normal, indicating that most countries have moderate savings rates, with a few exhibiting higher values. The majority of countries fall within a central range, with extremely high values being less common. Overall, the variability is moderate, suggesting some differences in savings patterns across nations. However, there are two outliers at the left, indicating negative amount of savings

Variable: Inflation

Original data file: WEO\_Oct2024.xlsx

Original variable name: Inflation, end of period consumer prices

Number of Observations: 558

No missing value

Definition: A consumer price index (CPI) measures changes in the prices of goods and services that households consume

Mean: 89.508

Standard Deviation: 31.17

Minimum: 35.211

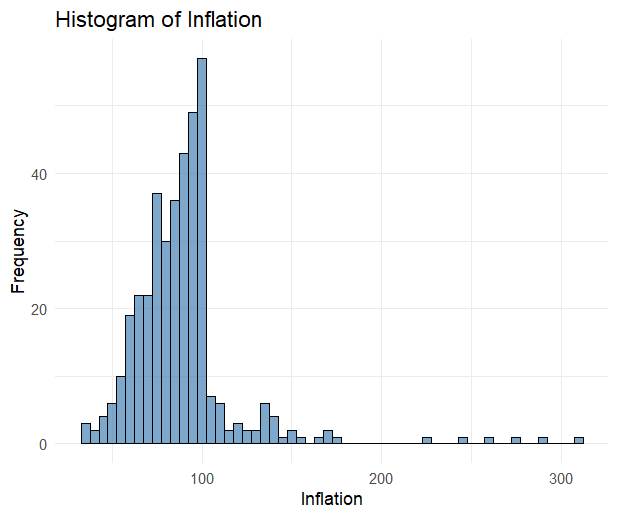
Median: 87.893

Maximum: 311.796

25th percentile: 74.523

75th percentile: 98.045

Histogram:



The distribution of inflation is right-skewed, with most values concentrated in a moderate range but a few extreme outliers (most of them are from Kazakhstan). The mean and median are relatively close, indicating a symmetric central tendency.

Variable: Unemployment rate

Original data file: WEO\_Oct2024.xlsx

Original variable name: Inflation, end of period consumer prices

Number of Observations: 558

No missing value

Definition: Unemployment rate gives the number of unemployed persons as a percentage of the labor force (the total number of people employed plus unemployed)

Mean: 8.780

Standard Deviation: 4.670

Minimum: 0.700

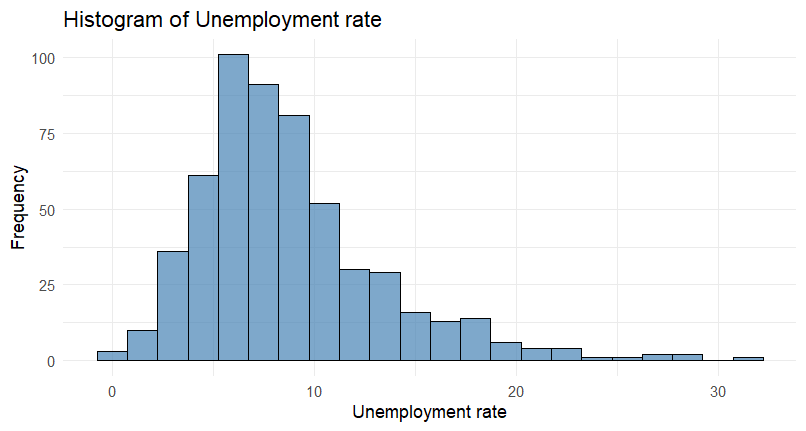
Median: 7.873

Maximum: 31.200

25th percentile: 5.747

75th percentile: 10.675

Histogram:



The distribution of the Unemployment Rate is right-skewed, with most values concentrated in the lower range but a few higher outliers. One counter-intuitive thing is that the outliers are not from countries with low incomes. Instead, they are from some high-income countries.

Variable: General government total expenditure

Original data file: WEO\_Oct2024.xlsx

Original variable name: General government revenue

Number of Observations: 558

No missing value

Definition: Revenue consists of taxes, social contributions, grants receivable, and other revenue as a percent of GDP

Mean: 37.806

Standard Deviation: 11.811

Minimum: 12.477

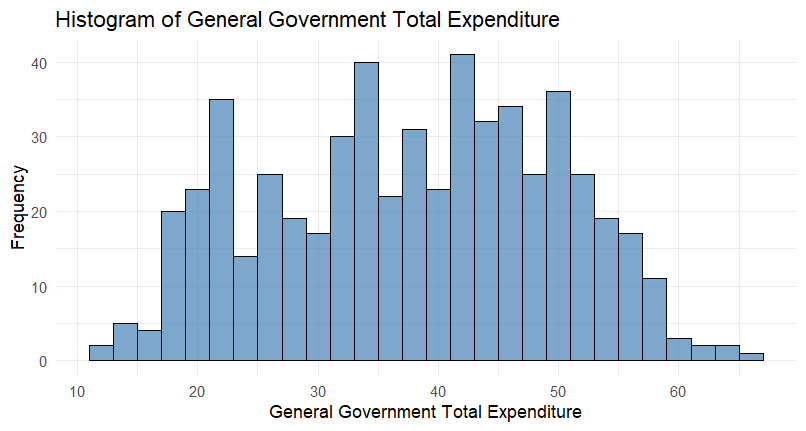
Median: 38.477

Maximum: 66.639

25th percentile: 28.454

75th percentile: 47.098

Histogram and box plot:



The distribution of General Government Revenue is approximately symmetric, with the mean and median being very close. Most values are concentrated within a moderate range, with no extreme outliers.

Variable: General government net lending borrowing

Original data file: WEO\_Oct2024.xlsx

Original variable name: General government net lending/borrowing

Number of Observations: 558

No missing value

Definition: Net lending (+)/ borrowing (-) is calculated as revenue minus total expenditure. as a percent of GDP

Mean: -2.471

Standard Deviation: 3.832

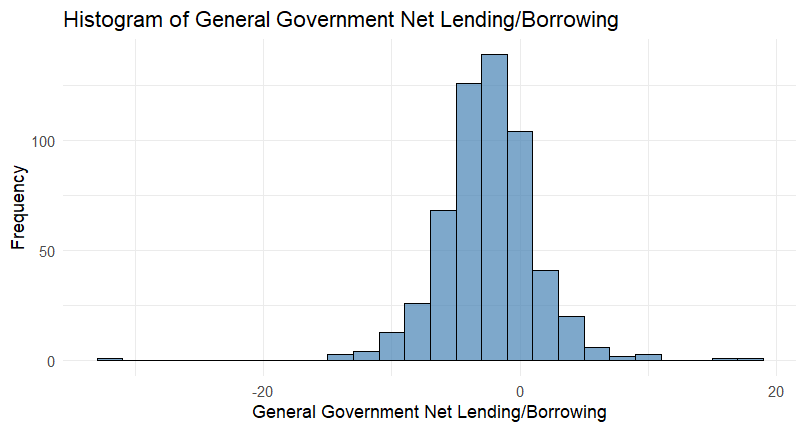
Minimum: -32.108

Median: -2.538

Maximum: 17.038

25th percentile: -4.436

75th percentile: -0.388

Histogram:

The distribution of General Government Net Lending/Borrowing (as percent of GDP) is left-skewed, with most values below zero and the interquartile range suggests that most countries have moderate deficits, with only a few approaching a balanced budget or a surplus.

Variable: Current account balance

Original data file: WEO\_Oct2024.xlsx

Original variable name: Current account balance

Number of Observations: 558

No missing value

Definition: The current account is all transactions other than those in financial and capital items. The major classifications are goods and services, income, and current transfers. As a percent of GDP

Mean: -2.130

Standard Deviation: 5.975

Minimum: -32.159

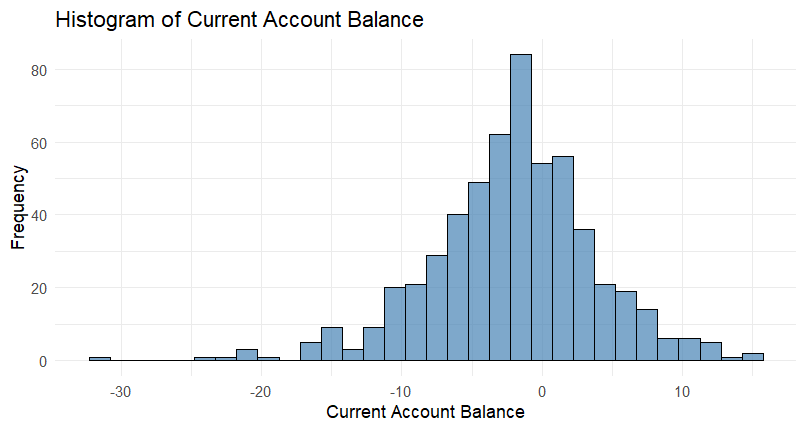
Median: -1.791

Maximum: 14.663

25th percentile: -5.321

75th percentile: 1.546

Histogram and box plot:



The Current Account Balance distribution is approximately normal, with a slight left skew. Most countries have a small deficit or surplus, with values concentrated near zero. The presence of both positive and negative outliers suggests that while some countries maintain significant surpluses, others experience substantial deficits.