Variable: Country Name, Country Code

Original data file: All data files

Original variable name: Country Name, Country Code

Number of Observations: 386

No missing value

Definition: as the name says, the “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 of using 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

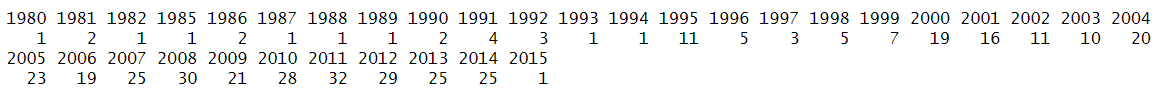
Original variable name: year (or Year)

Number of Observations: 386

No missing value

Definition: The year of the observation

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

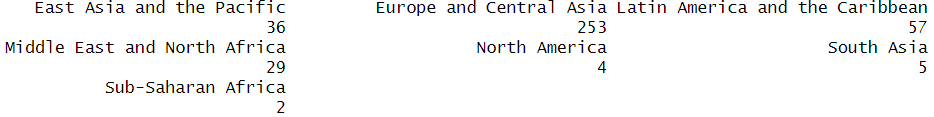
Original variable name: region\_wb

Number of Observations: 386

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: 386

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: 386

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: 386

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.

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: 386

No missing value

Definition: Personal remittances, received in % of GDP

Mean:1.624

Standard Deviation:2.477

Minimum:0

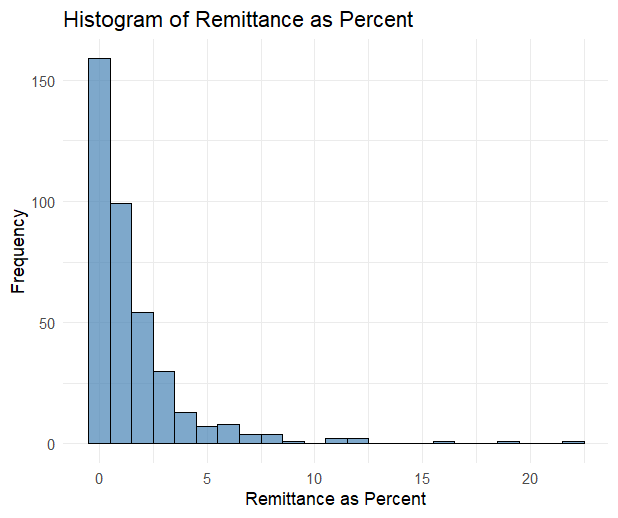
Median: 0.744

Maximum: 21.589

25th percentile:0.258

75th percentile:2.086

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: 386

No missing value

Definition: Personal remittances, received (current US$)

Mean: 3145192441

Standard Deviation:5097721081

Minimum:0

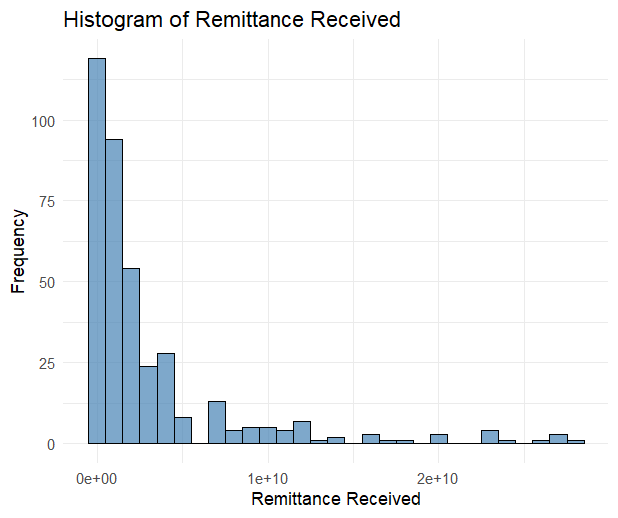
Median: 1105917678

Maximum: 27976008592

25th percentile: 414513324.6

75th percentile: 3310012285

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: 386

No missing value

Definition: Personal remittances, paid (current US$)

Mean: 3145192441

Standard Deviation: 4079310704

Minimum:0

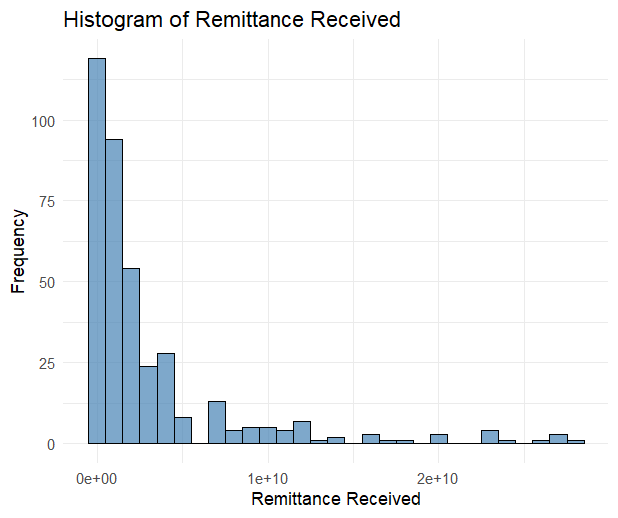
Median: 622279774.9

Maximum: 26816169395

25th percentile: 179325388.2

75th percentile: 2561219674

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: 386

No missing value

Definition: Standardized Gini coefficient

Mean: 36.172

Standard Deviation: 8.809

Minimum: 22.19

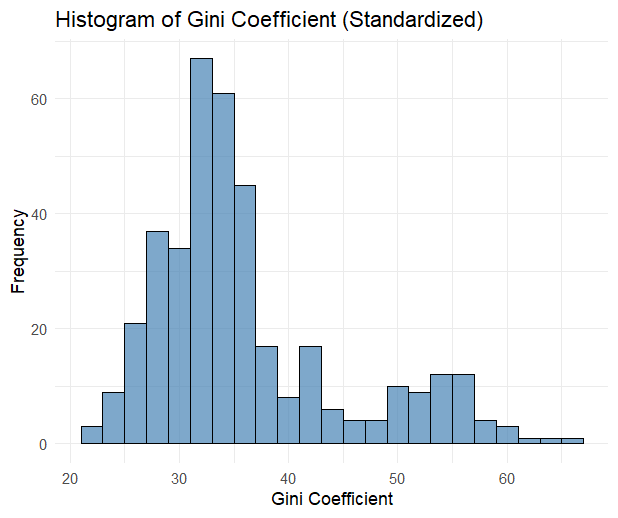
Median: 33.76

Maximum: 66.66

25th percentile: 30.63

75th percentile: 38.568

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: 386

No missing value

Definition: Income share of the bottom 20%

Mean: 6.823

Standard Deviation: 2.073

Minimum: 1.008

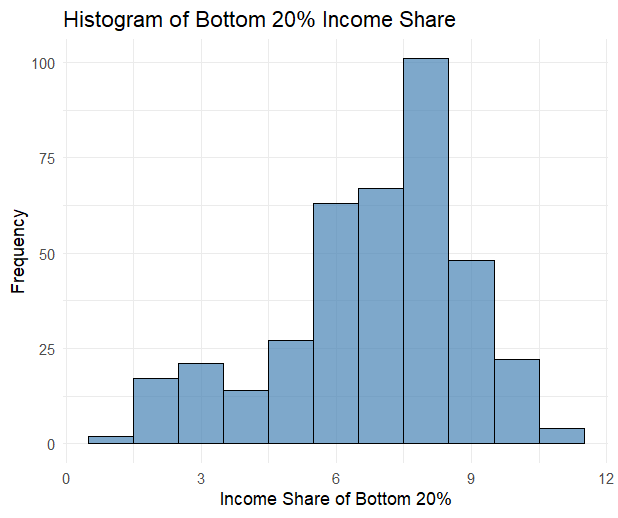
Median: 7.2

Maximum: 10.8

25th percentile: 5.8

75th percentile: 8.239

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 some notable outliers close to 0, representing extreme income inequality. Interestingly, these outliers are largely the same as those observed in the Gini coefficient distribution, reinforcing the pattern of severe inequality in certain countries.

Variable: top20

Original data file: wiidcountry.xlsx

Original variable name: top20

Number of Observations: 386

No missing value

Definition: Income share of the top 20%

Mean: 36.491

Standard Deviation: 3.728

Minimum: 30.284

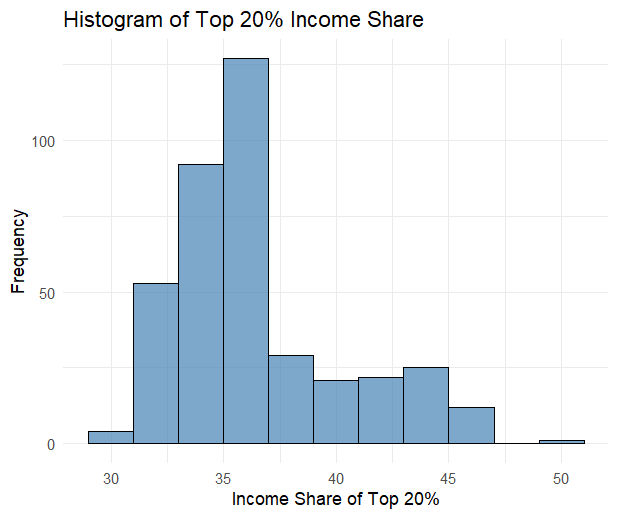
Median: 35.59

Maximum: 50.291

25th percentile: 33.955

75th percentile: 38.027

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: 386

No missing value

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

Mean: 12.343

Standard Deviation: 1.343

Minimum: 6.254

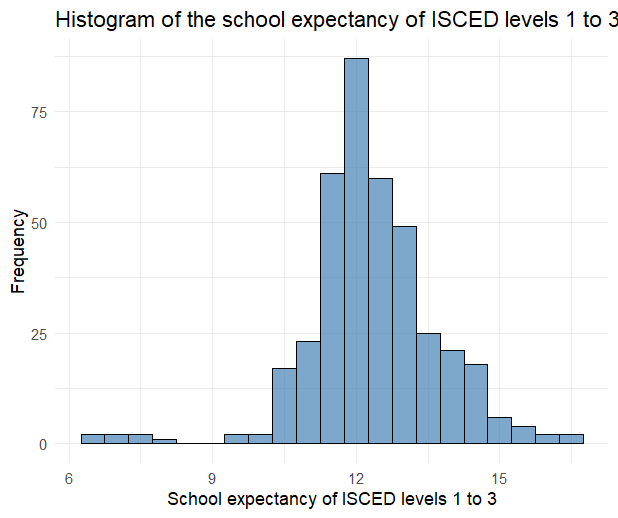
Median:12.172

Maximum: 16.462

25th percentile: 11.696

75th percentile: 13.104

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: 386

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: 2.715

Standard Deviation: 0.898

Minimum: 0.158

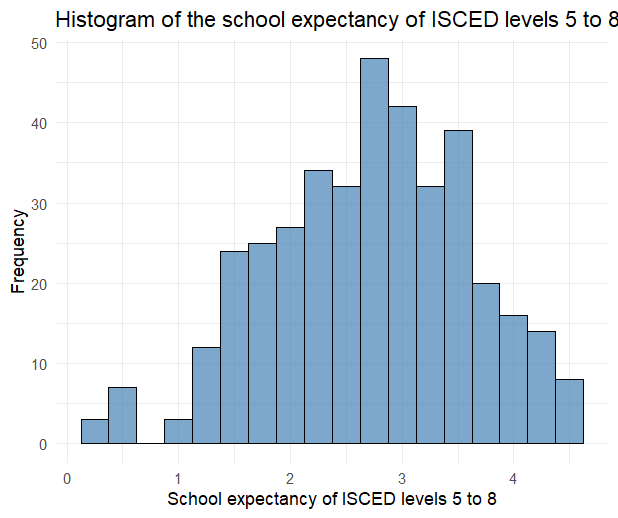
Median: 2.739

Maximum: 4.608

25th percentile: 2.093

75th percentile: 3.377

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 almost no outlier other than some values that are very close to 0.

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: 386

No missing value

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

Mean: 641.802

Standard Deviation: 876.329

Minimum: 2.148

Median: 305.092

Maximum:4581.459

25th percentile: 98.093

75th percentile: 588.699

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: 386

No missing value

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

Mean: 36972.916

Standard Deviation: 18899.883

Minimum: 4720.947

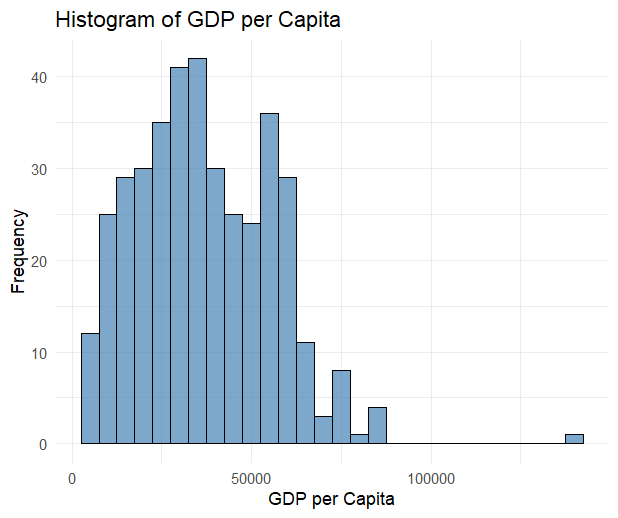
Median: 35579.844

Maximum: 138915.885

25th percentile: 22632.52

75th percentile: 52156.634

Histogram and Boxplot:





In general, the distribution of GDO per Capita is relatively symmetric and close to normal distribution. The standard deviation is relatively low. However, this is an outlier, which is Luxembourg 2008.

Variable: Total investment

Original data file: WEO\_Oct2024.xlsx

Original variable name: Total investment

Number of Observations: 386

No missing value

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

Mean: 23.614

Standard Deviation: 4.86

Minimum: 12.676

Median: 22.87

Maximum: 43.351

25th percentile: 20.715

75th percentile: 25.534

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, this shows that most observations have value around these two numbers.