INFO212 Assignment2

GROUP 22

Name: Qinru Yang ID: 320220940961

Name: Rongshen He ID: 320220940271

Name: Qirui Zhang ID: 320220941080

Name: Wei Dou ID: 320220940190

Catalogue

Step1: Data Collection

First, load the common library required for this assignment.

The data we use is from **the World Bank**. We first call the World Bank API to query data on GDP, unemployment rate, inflation, and other metrics for five countries, spanning nearly 40 years. We use the API because most of the existing datasets available online are not in JSON format. Since JSON data is often used for efficient front-end and back-end transmission and contains some unnecessary parameters, we decided to use the API to obtain the dataset in JSON format.

We constructs a URL for each country and indicator, sends a request to the World Bank API, and processes the JSON response. The data for each indicator and country is stored in a dictionary.Finally, the combined data is saved into a JSON file named "**world_bank_data.json**".

Step2: Data Preparation

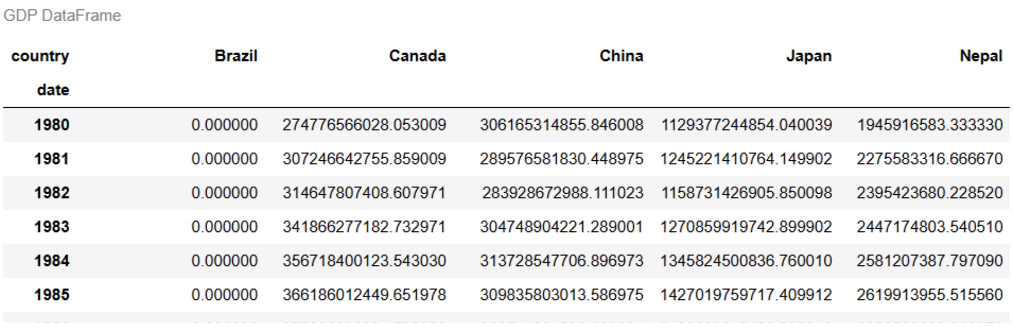
To convert the data in the json file into dataframes and perform data cleaning, we designed three functions.

The first function "**create_indicator_dataframe**": it transforms raw JSON data for an economic indicator into a pivoted pandas DataFrame. It extracts country names, years, and values, creating a structured DataFrame with years as rows and countries as columns, facilitating easy analysis.

The second function "**display_dfs**": it walks through the generated data boxes converts them to a more observable HTML format.

The third function "**clear_Nan function**": it clears the nan data in the DataFrame. It creates a **mask** indicating the locations of NaN values, replaces them with the specified value, and then returns the mask.

So we get the cleaned data as follow (one of the indicators):



GDP DataFrame

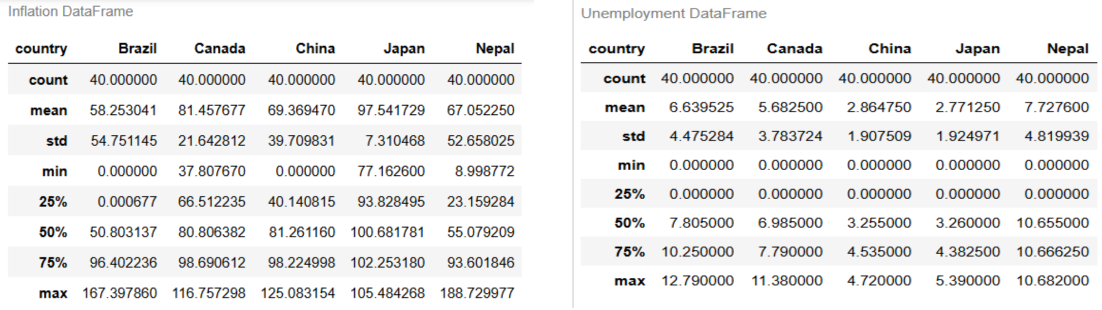
country	Brazil	Canada	China	Japan	Nepal
date					
1980	0.000000	274776566028.053009	306165314855.846008	1129377244854.040039	1945916583.333330
1981	0.000000	307246642755.859009	289576581830.448975	1245221410764.149902	2275583316.666670
1982	0.000000	314647807408.607971	283928672988.111023	1158731426905.850098	2395423680.228520
1983	0.000000	341866277182.732971	304748904221.289001	1270859919742.899902	2447174803.540510
1984	0.000000	356718400123.543030	313728547706.896973	1345824500836.760010	2581207387.797090
1985	0.000000	366186012449.651978	309835803013.586975	1427019759717.409912	2619913955.515560

Table 1

Step3: Data Analysis

We perform exploratory data analysis (**EDA**) to understand the datasets. To get a summary of key statistics (mean, standard deviation, min, max) for each indicator DataFrame. We directly call the pandas function **describe()** to show the data distribution of each indicator.

The results of two of the indicator tables are shown in table2:



Inflation DataFrame

country	Brazil	Canada	China	Japan	Nepal
count	40.000000	40.000000	40.000000	40.000000	40.000000
mean	58.253041	81.457677	69.369470	97.541729	67.052250
std	54.751145	21.642812	39.709831	7.310468	52.658025
min	0.000000	37.807670	0.000000	77.162600	8.998772
25%	0.000677	66.512235	40.140815	93.828495	23.159284
50%	50.803137	80.806382	81.261160	100.681781	55.079209
75%	96.402236	98.690612	98.224998	102.253180	93.601846
max	167.397860	116.757298	125.083154	105.484268	188.729977

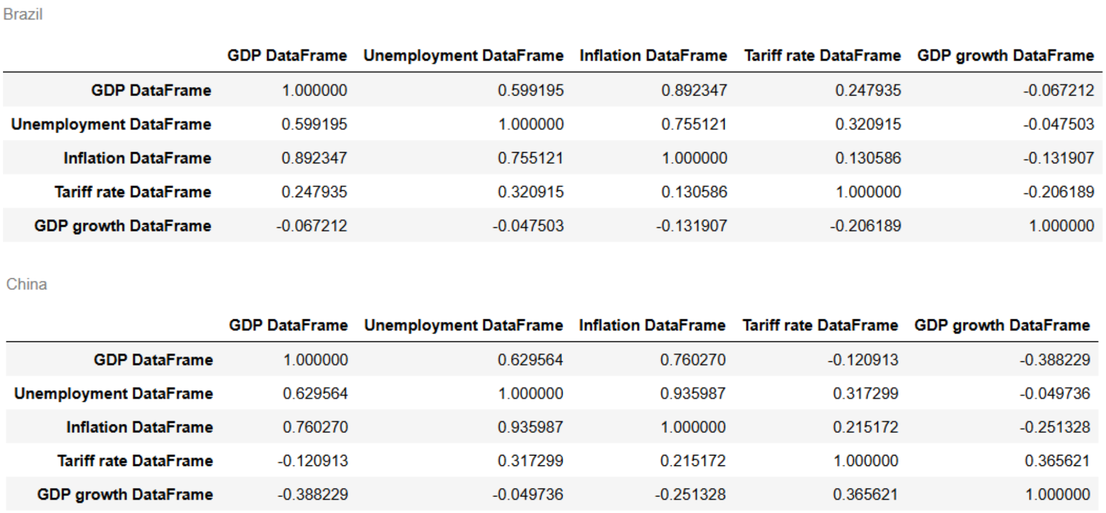
Unemployment DataFrame

country	Brazil	Canada	China	Japan	Nepal
count	40.000000	40.000000	40.000000	40.000000	40.000000
mean	6.639525	5.682500	2.864750	2.771250	7.727600
std	4.475284	3.783724	1.907509	1.924971	4.819939
min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000	0.000000	0.000000
50%	7.805000	6.985000	3.255000	3.260000	10.655000
75%	10.250000	7.790000	4.535000	4.382500	10.666250
max	12.790000	11.380000	4.720000	5.390000	10.682000

Table 2

In order to correlations or anomalies between different economic indicators, we calculated **matrix of correlation coefficients**. We traverse the dataframes and using the **df.corr()** function to calculate the correlation between the indicators.

The matrix of correlation coefficients for two of the countries is shown table3:



Brazil

	GDP DataFrame	Unemployment DataFrame	Inflation DataFrame	Tariff rate DataFrame	GDP growth DataFrame
GDP DataFrame	1.000000	0.599195	0.892347	0.247935	-0.067212
Unemployment DataFrame	0.599195	1.000000	0.755121	0.320915	-0.047503
Inflation DataFrame	0.892347	0.755121	1.000000	0.130586	-0.131907
Tariff rate DataFrame	0.247935	0.320915	0.130586	1.000000	-0.206189
GDP growth DataFrame	-0.067212	-0.047503	-0.131907	-0.206189	1.000000

China

	GDP DataFrame	Unemployment DataFrame	Inflation DataFrame	Tariff rate DataFrame	GDP growth DataFrame
GDP DataFrame	1.000000	0.629564	0.760270	-0.120913	-0.388229
Unemployment DataFrame	0.629564	1.000000	0.935987	0.317299	-0.049736
Inflation DataFrame	0.760270	0.935987	1.000000	0.215172	-0.251328
Tariff rate DataFrame	-0.120913	0.317299	0.215172	1.000000	0.365621
GDP growth DataFrame	-0.388229	-0.049736	-0.251328	0.365621	1.000000

Table 3

What's more, the trends over time will be presented in the line charts and bar charts in the next step.

Step4: Data Visualization

Bar chart of GDP over time for five countries (Figure1). Our codes creates a bar chart based on the dataframe to visualize GDP data for five countries over time. The chart uses a different color for each country and includes an interactive feature that enhances exploration. In addition, the bar chart is overlaid with a red line representing **the average GDP growth of all countries**, highlighting the overall trend. **This comprehensive visualization provides a clear comparison of GDP growth by country and year**, as well as an overall trend line of average GDP growth.

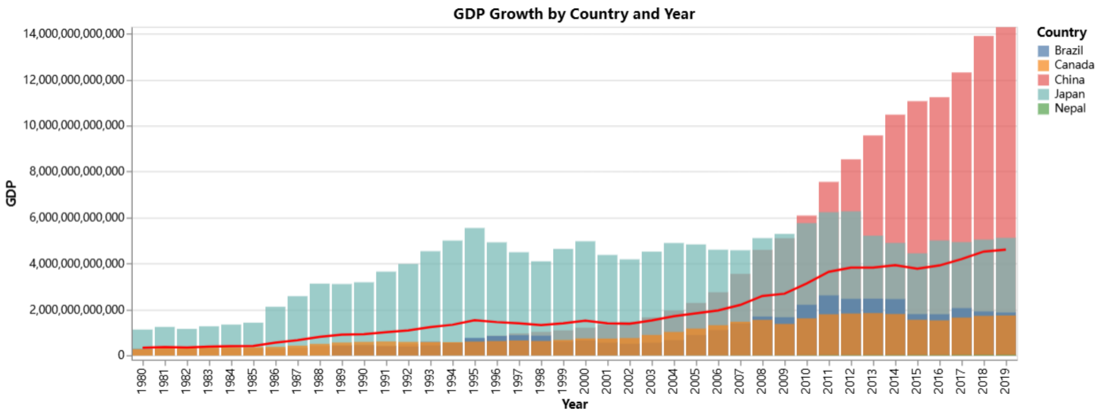


Figure 1

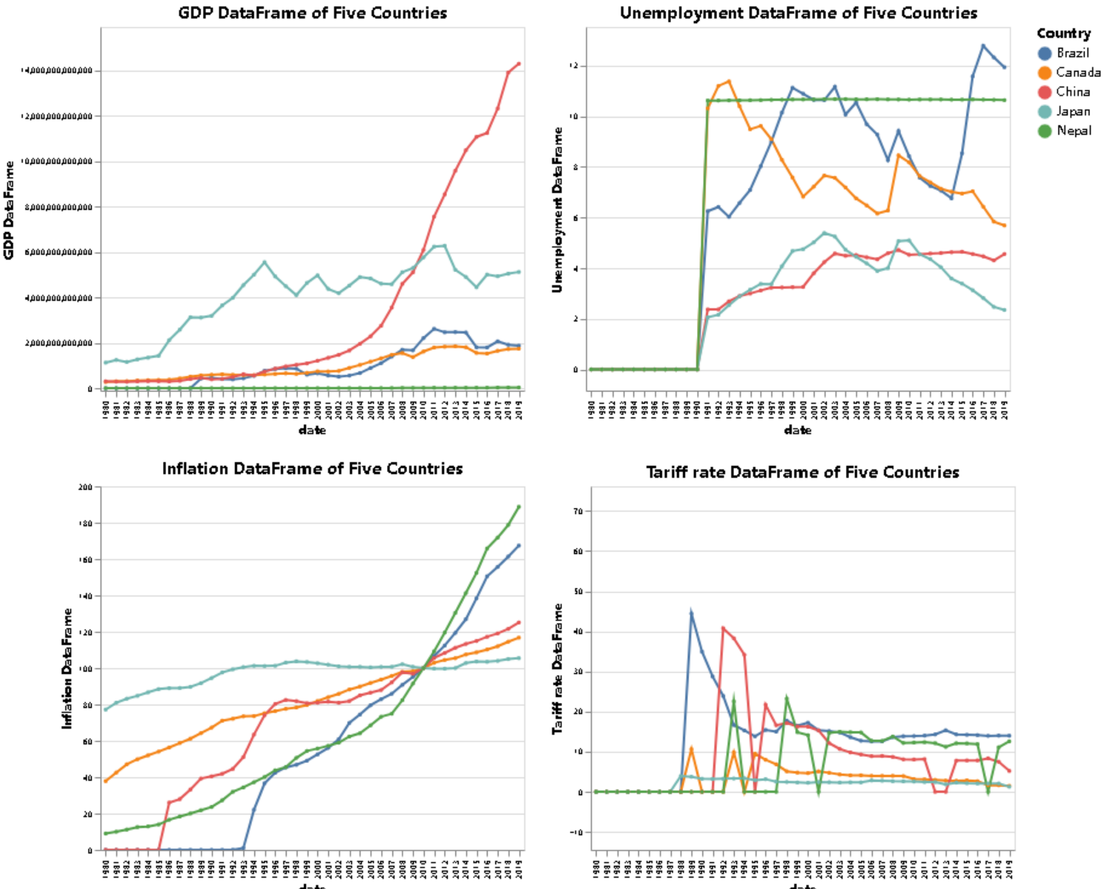


Figure 2

Comparison of line charts for different indicators in different countries (Figure2). We defined a function “draw_line_chart” to create an interactive line chart for each indicator in the five countries. **Each chart shows trends over time, with different colors representing different countries.**

Pair plots (Figure3). We generates pair plots using Seaborn to visualize the relationships between GDPs of different countries. Each pair plot shows how the indicator of one country correlates with the indicators of other countries, highlighting patterns and potential influences between them. One of the pair plots is as figure3:

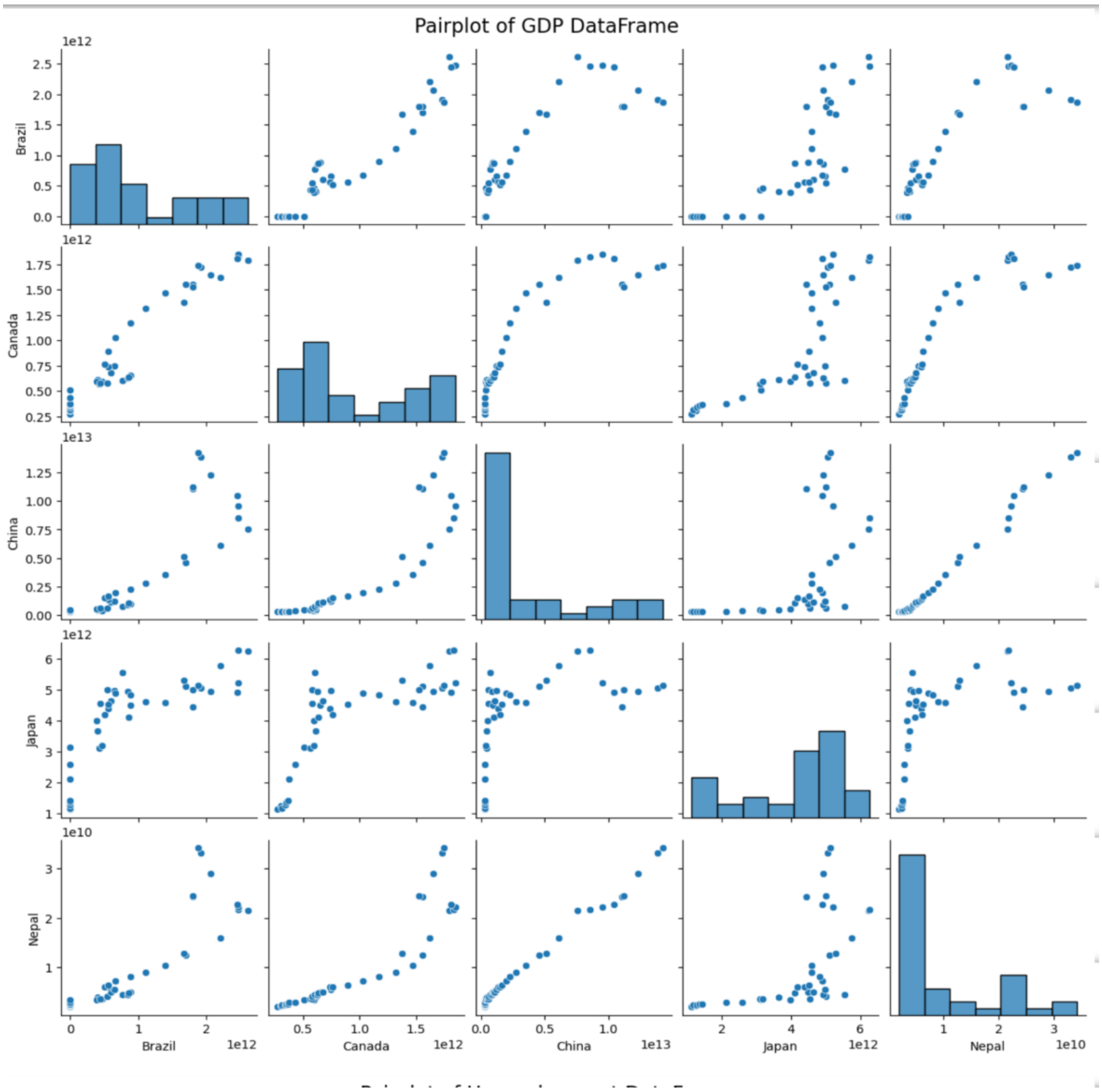


Figure 3

Heat maps by country (Figure4). Using the **Seaborn** library, visualize correlations between individual data for each country in the df_corrs dictionary to create heat maps. Each heat map uses color gradients to show the strength and direction of the relationship between the indicators. In addition, the average heat map is calculated by taking the average of all the relevant matrices to provide a comprehensive view of the overall trends and patterns of the data for all countries. The average heat map is shown

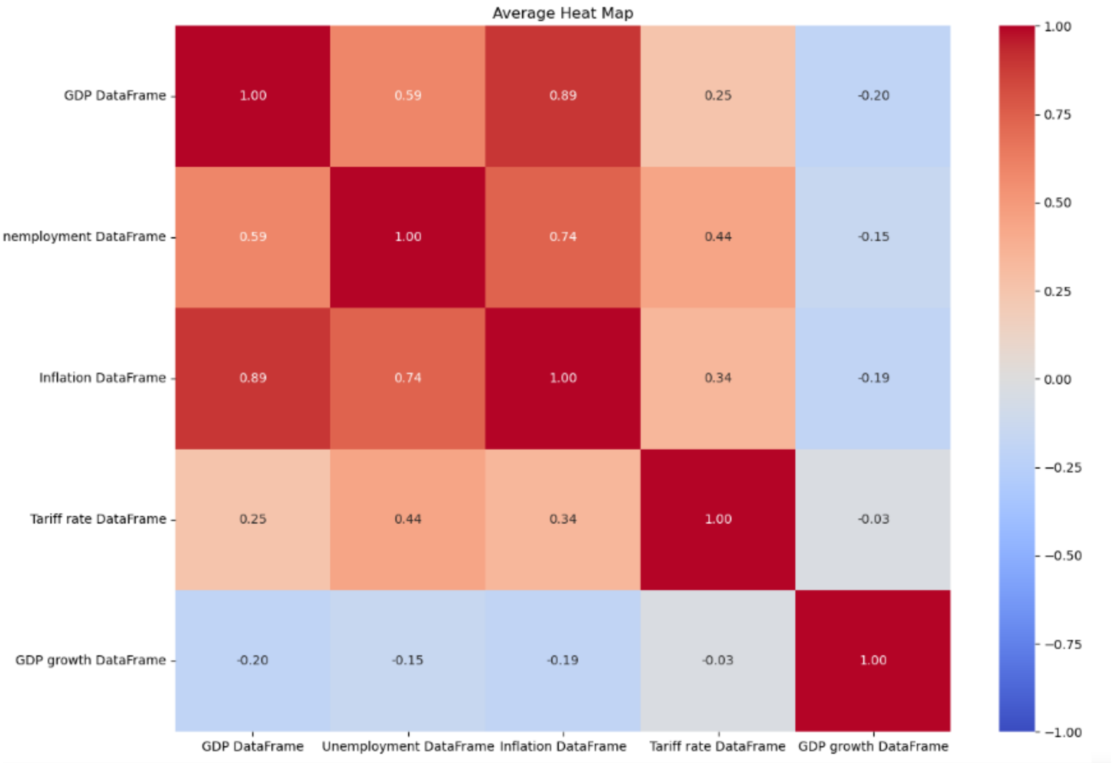


Figure 4

in figure 4 (for other heat maps, please check our notebook):

Box plots of GDP growth (Fiugure5). We create box plots for the GDP growth rates of different countries, allowing a quick visual comparison of their economic performance. Box plots summarize key statistics like the median, quartiles, and outliers, making it easy to identify trends, variability, and unusual data points across countries. It simplifies complex data, helping to highlight differences and similarities in GDP growth rates effectively. Two of the box plots are as figure5:

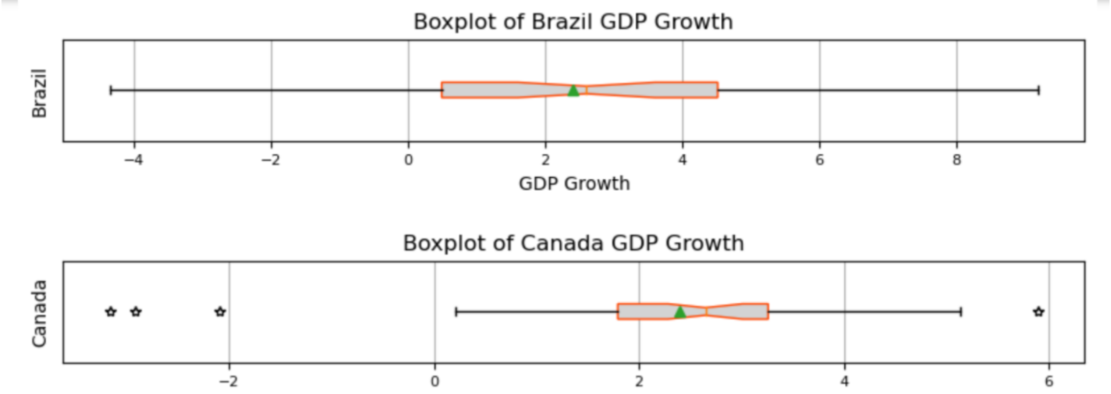


Figure 5

Note: There are a lot of plots generated in our jupyter notebook, for reasons of space, we only show representative images in the report (other plots not shown are similar). To see the full figure, check out our jupyter notebook.

Step5: Interpretation

Analysis GDP Growth by Country and Year Picture

GDP growth: The average level of global GDP has risen steadily over the past 40 years, showing the continued improvement of the world economy in recent years.

Japan is the best economy until 2010, and a clear lead until 1991 in the 5 countries. But China urpassed Japan after 2010, and the economic gap widened with each passing year. As the two economic powers in Asia, their economic trends show that Japan's economic development lags behind China's after the 21st century. East Asia's economic centre of gravity is shifting towards China.

Analysis Line chart of different indicators

Unemployment rate: High from 1999 to 2004, then declining steadily.

Inflation: Rising since 1980, Nepal has risen most sharply, possibly due to internal unrest and an unstable external environment.

Tariff rates: have declined gradually since 1980, reflecting continued openness and progress in world trade.

GDP growth rate: The GDP growth rate of each country fluctuates around a specific average, especially in China.

Analysis pair plots

Cross-country indicator correlation: Most indicators are not correlated.

GDP and tariffs: Strong positive correlation, reflecting trade cooperation and co-growth between countries, such as frequent or co-rapid growth between China and Nepal.

Tariffs and unemployment: Positive correlation, likely reflecting declining domestic economic conditions leading to higher unemployment and higher tariffs.

Analysis heatmaps

Correlation between indicators: There is little difference in correlation between indicators among countries, reflecting the same trend.

GDP and inflation rate: Closely correlated, showing that GDP does not directly reflect national prosperity.

High unemployment and inflation: closely related and can occur simultaneously in poor economic conditions.

High unemployment and total GDP: The correlation is about 0.6, indicating that higher unemployment can lead to higher GDP, which is difficult to explain.

Analysis box-plot picture

Differences in GDP growth rates: GDP growth rates vary significantly among countries, with China significantly higher than other countries.

Negative growth outliers: Japan, Canada and Nepal have negative growth outliers, which may be related to the economic crisis or pandemic crisis.

Analysis Correlations

GDP growth typically inversely correlates with unemployment

Inflation and unemployment have a short-term inverse relationship , but it's unstable long-term.

GDP and tariffs may correlate positively in countries heavily reliant on tariff revenues, though high tariffs can long-term hinder trade and GDP growth.

GDP growth can correlate with inflation during demand-pushed price increases, but high inflation can stifle further growth.

Analysis Abnormal

The heatmap shows a correlation of approximately 0.6 between high unemployment and total GDP, suggesting that higher unemployment might be associated with higher GDP, which is challenging to explain.

The pair chart indicates that tariff rates between countries do not exhibit a strong correlation. However, in general, tariff rates can influence each other.

The inflation rates of five countries were all 100 in 2011, which seems highly coincidental. However, the data verified on the official website is correct, and no errors were found in data processing.

The unemployment rate in Nepal has consistently been around 10.6% for the past 30 years, which seems unusual. If the data is accurate, it reflects a long-term severe employment situation in Nepal.