G20 Countries Economic Indicators

Dataset description

Country Name: The name of the country being analyzed.

Country Code: The unique code for each country (ISO Alpha-3 code).

Time: The year of the data point.

Time Code: The code for the time period (usually corresponds to the year).

Current account balance (% of GDP) [BN.CAB.XOKA.GD.ZS]: The balance of a country's current account, including trade balance, primary income, and secondary income, as a percentage of GDP.

Exports of goods and services (% of GDP) [NE.EXP.GNFS.ZS]: The value of all goods and services sold abroad, expressed as a percentage of GDP.

Foreign direct investment, net inflows (% of GDP) [BX.KLT.DINV.WD.GD.ZS]: The inflow of foreign investment into a country, as a percentage of GDP.

Foreign direct investment, net outflows (% of GDP) [BM.KLT.DINV.WD.GD.ZS]: The outflow of investment from a country to foreign countries, as a percentage of GDP.

GDP (current US$) [NY.GDP.MKTP.CD]: The total monetary value of goods and services produced in a country, in current US dollars.

GDP growth (annual %) [NY.GDP.MKTP.KD.ZG]: The annual growth rate of GDP.

GDP per capita (current US$) [NY.GDP.PCAP.CD]: GDP divided by the midyear population, indicating the economic output per person.

GDP per capita growth (annual %) [NY.GDP.PCAP.KD.ZG]: The annual growth rate of GDP per capita.

General government final consumption expenditure (% of GDP) [NE.CON.GOVT.ZS]: Total government spending as a percentage of GDP.

Government expenditure on education, total (% of GDP) [SE.XPD.TOTL.GD.ZS]: Total government spending on education as a percentage of GDP.

Government expenditure per student, primary (% of GDP per capita) [SE.XPD.PRIM.PC.ZS]: Government spending per primary school student as a percentage of GDP per capita.

Government expenditure per student, secondary (% of GDP per capita) [SE.XPD.SECO.PC.ZS]: Government spending per secondary school student as a percentage of GDP per capita.

Government expenditure per student, tertiary (% of GDP per capita) [SE.XPD.TERT.PC.ZS]: Government spending per tertiary-level student as a percentage of GDP per capita.

Gross savings (% of GDP) [NY.GNS.ICTR.ZS]: Total savings as a percentage of GDP, reflecting income saved rather than spent.

Imports of goods and services (% of GDP) [NE.IMP.GNFS.ZS]: The value of all goods and services purchased from abroad, expressed as a percentage of GDP.

GNI (current US$) [NY.GNP.MKTP.CD]: Gross national income in current US dollars.

GNI growth (annual %) [NY.GNP.MKTP.KD.ZG]: The annual growth rate of GNI.

Employment and Labor

Employment to population ratio, 15+, total (%) (national estimate) [SL.EMP.TOTL.SP.NE.ZS]: The ratio of employed individuals aged 15 and older to the total working-age population.

Employment to population ratio, 15+, female (%) (national estimate) [SL.EMP.TOTL.SP.FE.NE.ZS]: The ratio of employed females aged 15 and older to the total working-age female population.

Employment to population ratio, 15+, male (%) (national estimate) [SL.EMP.TOTL.SP.MA.NE.ZS]: The ratio of employed males aged 15 and older to the total working-age male population.

Unemployment, total (% of total labor force) (national estimate) [SL.UEM.TOTL.NE.ZS]: The percentage of the labor force that is unemployed.

Employment in services (% of total employment) (modeled ILO estimate) [SL.SRV.EMPL.ZS]: The percentage of employed persons working in the services sector.

Employment in industry (% of total employment) (modeled ILO estimate) [SL.IND.EMPL.ZS]: The percentage of employed persons working in the industrial sector.

Employment in agriculture (% of total employment) (modeled ILO estimate) [SL.AGR.EMPL.ZS]: The percentage of employed persons working in agriculture.

Poverty headcount ratio at national poverty lines (% of population) [SI.POV.NAHC]: The percentage of the population living below the national poverty line.

Multidimensional poverty headcount ratio (UNDP) (% of population) [SI.POV.MPUN]: The percentage of the population experiencing multiple deprivations (education, health, and living standards).

Fixed broadband subscriptions [IT.NET.BBND]: The number of fixed broadband subscriptions.

Fixed telephone subscriptions [IT.MLT.MAIN]: The number of fixed telephone subscriptions.

Mobile cellular subscriptions [IT.CEL.SETS]: The number of mobile cellular subscriptions.

Secure Internet servers [IT.NET.SECR]: The number of secure servers that encrypt web communications.

High-technology exports (current US$) [TX.VAL.TECH.CD]: The value of high-tech exports, in current US dollars.

High-technology exports (% of manufactured exports) [TX.VAL.TECH.MF.ZS]: The percentage of manufactured exports that are high-technology products.

Research and development expenditure (% of GDP) [GB.XPD.RSDV.GD.ZS]: R&D expenditure as a percentage of GDP.

Scientific and technical journal articles [IP.JRN.ARTC.SC]: The number of scientific and technical articles published.

Air transport, freight (million ton-km) [IS.AIR.GOOD.MT.K1]: Freight transported by air, in million ton-km.

Air transport, passengers carried [IS.AIR.PSGR]: The total number of passengers carried by air.

Rail lines (total route-km) [IS.RRS.TOTL.KM]: The total length of rail routes, in kilometers.

Railways, goods transported (million ton-km) [IS.RRS.GOOD.MT.K6]: Goods transported by rail, in million ton-km.

Railways, passengers carried (million passenger-km) [IS.RRS.PASG.KM]: Passenger travel by rail, in million passenger-km.

Community health workers (per 1,000 people) [SH.MED.CMHW.P3]: The number of community health workers per 1,000 people.

Current health expenditure (% of GDP) [SH.XPD.CHEX.GD.ZS]: Health expenditure as a percentage of GDP.

Current health expenditure per capita (current US$) [SH.XPD.CHEX.PC.CD]: Health expenditure per person, in current US dollars.

Hospital beds (per 1,000 people) [SH.MED.BEDS.ZS]: The number of hospital beds per 1,000 people.

Physicians (per 1,000 people) [SH.MED.PHYS.ZS]: The number of physicians per 1,000 people.

Life expectancy at birth, total (years) [SP.DYN.LE00.IN]: The average number of years a newborn is expected to live.

Probability of dying among adolescents ages 10-14 years (per 1,000) [SH.DYN.1014]: The probability of death per 1,000 children aged 10-14.

Probability of dying among adolescents ages 15-19 years (per 1,000) [SH.DYN.1519]: The probability of death per 1,000 teenagers aged 15-19.

Probability of dying among children ages 5-9 years (per 1,000) [SH.DYN.0509]: The probability of death per 1,000 children aged 5-9.

Probability of dying among youth ages 20-24 years (per 1,000) [SH.DYN.2024]: The probability of death per 1,000 youth aged 20-24.

Survival to age 65, female (% of cohort) [SP.DYN.TO65.FE.ZS]: The percentage of females expected to survive to age 65.

Survival to age 65, male (% of cohort) [SP.DYN.TO65.MA.ZS]: The percentage of males expected to survive to age 65.

This is a Jupyter notebook with the extension .ipynb, the python version being used in this project is 3.9.12.

In order to run the cells just click [Shift + Enter] or [Ctrl + Enter]

In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
```

In [2]:
```
sns.set_theme(style="whitegrid")
```

In [3]:
```
df = pd.read_csv("WorldEconomicIndicators.csv")
df.head()
```

Out[3]:

	Country Name	Country Code	Time	Time Code	Current account balance (% of GDP)	Exports of goods and services (% of GDP)	Foreign direct investment, net inflows (% of GDP)	Foreign direct investment, net outflows (% of GDP)	GDP (current US$)	GDP growth (annual %)	...	Current health expenditure per capita (current US$)	Hospital beds (per 1,000 people)	Physicians (per 1,000 people)	exp
0	Pakistan	PAK	2000	YR2000	-0.085440186	9.629626104	0.309595026	0.011056965	99484802345	4.260088011	...	15.1975832	0.5	0.595	
1	Pakistan	PAK	2001	YR2001	1.933180342	10.62617308	0.389106586	0.026763945	97145618480	3.651350171	...	13.20521545	0.5	0.605	
2	Pakistan	PAK	2002	YR2002	3.935733262	11.26296557	0.8435173	0.031657429	97923302809	2.594816684	...	14.74764633	0.49	..	
3	Pakistan	PAK	2003	YR2003	3.179620139	12.16464391	0.475207712	0.01690814	1.12E+11	5.401310873	...	14.1153574	0.48	..	
4	Pakistan	PAK	2004	YR2004	-0.617928013	11.42297416	0.845585702	0.042354919	1.32E+11	7.83125557	...	15.29145527	0.48	0.682	

5 rows × 55 columns

In [4]:
```
df.shape
```

Out[4]:
```
(1564, 55)
```

We do not need the Time Code column, so we will remove it

In [5]:
```
try:
    df.drop("Time Code", axis=1, inplace=True)
    print("Column dropped")
except Exception as e:
    print(e)
```

```
Column dropped
```

In [6]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1564 entries, 0 to 1563
Data columns (total 54 columns):
 #   Column                                                                Non-Null Count  Dtype
---  ------                                                                --------------  -----
 0   Country Name                                                          1564 non-null   object
 1   Country Code                                                          1564 non-null   object
 2   Time                                                                  1564 non-null   int64
 3   Current account balance (% of GDP)                                    1564 non-null   object
 4   Exports of goods and services (% of GDP)                              1564 non-null   object
 5   Foreign direct investment, net inflows (% of GDP)                     1564 non-null   object
 6   Foreign direct investment, net outflows (% of GDP)                    1564 non-null   object
 7   GDP (current US$)                                                     1564 non-null   object
 8   GDP growth (annual %)                                                 1564 non-null   object
 9   GDP per capita (current US$)                                          1564 non-null   object
 10  GDP per capita growth (annual %)                                      1564 non-null   object
 11  General government final consumption expenditure (% of GDP)           1564 non-null   object
 12  Government expenditure on education, total (% of GDP)                 1564 non-null   object
 13  Government expenditure per student, primary (% of GDP per capita)     1564 non-null   object
 14  Government expenditure per student, secondary (% of GDP per capita)   1564 non-null   object
 15  Government expenditure per student, tertiary (% of GDP per capita)    1564 non-null   object
 16  Gross savings (% of GDP)                                              1564 non-null   object
 17  Imports of goods and services (% of GDP)                              1564 non-null   object
 18  GNI (current US$)                                                     1564 non-null   object
 19  GNI growth (annual %)                                                 1564 non-null   object
 20  Employment to population ratio, 15+, total (%) (national estimate)    1564 non-null   object
 21  Employment to population ratio, 15+, female (%) (national estimate)   1564 non-null   object
 22  Employment to population ratio, 15+, male (%) (national estimate)     1564 non-null   object
 23  Unemployment, total (% of total labor force) (national estimate)      1564 non-null   object
 24  Employment in services (% of total employment) (modeled ILO estimate) 1564 non-null   object
 25  Employment in industry (% of total employment) (modeled ILO estimate) 1564 non-null   object
 26  Employment in agriculture (% of total employment) (modeled ILO estimate) 1564 non-null   object
 27  Poverty headcount ratio at national poverty lines (% of population)   1564 non-null   object
 28  Multidimensional poverty headcount ratio (UNDP) (% of population)     1564 non-null   object
 29  Fixed broadband subscriptions                                         1564 non-null   object
 30  Fixed telephone subscriptions                                         1564 non-null   object
 31  Mobile cellular subscriptions                                         1564 non-null   object
 32  Secure Internet servers                                               1564 non-null   object
 33  High-technology exports (current US$)                                 1564 non-null   object
 34  High-technology exports (% of manufactured exports)                   1564 non-null   object
 35  Research and development expenditure (% of GDP)                       1564 non-null   object
 36  Scientific and technical journal articles                             1564 non-null   object
 37  Air transport, freight (million ton-km)                               1564 non-null   object
 38  Air transport, passengers carried                                     1564 non-null   object
 39  Rail lines (total route-km)                                           1564 non-null   object
 40  Railways, goods transported (million ton-km)                          1564 non-null   object
 41  Railways, passengers carried (million passenger-km)                   1564 non-null   object
 42  Community health workers (per 1,000 people)                          1564 non-null   object
 43  Current health expenditure (% of GDP)                                 1564 non-null   object
 44  Current health expenditure per capita (current US$)                   1564 non-null   object
 45  Hospital beds (per 1,000 people)                                      1564 non-null   object
 46  Physicians (per 1,000 people)                                         1564 non-null   object
 47  Life expectancy at birth, total (years)                               1564 non-null   float64
 48  Probability of dying among adolescents ages 10-14 years (per 1,000)   1564 non-null   object
 49  Probability of dying among adolescents ages 15-19 years (per 1,000)   1564 non-null   object
 50  Probability of dying among children ages 5-9 years (per 1,000)        1564 non-null   object
 51  Probability of dying among youth ages 20-24 years (per 1,000)         1564 non-null   object
 52  Survival to age 65, female (% of cohort)                              1564 non-null   float64
 53  Survival to age 65, male (% of cohort)                                1564 non-null   float64
dtypes: float64(3), int64(1), object(50)
memory usage: 659.9+ KB
```

Many of the columns which should have been numerical are float, so we will convert them to float

In [7]:
```
for i in df.columns[3:]: #the first three columns have the correct Data type
    try:
        df[i] = df[i].astype("float64")
        print(f"Data type for {i} changed successfully")
    except Exception as e:
        print(e)
```

```
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
Data type for Life expectancy at birth, total (years) changed successfully
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
could not convert string to float: '..'
Data type for Survival to age 65, female (% of cohort) changed successfully
Data type for Survival to age 65, male (% of cohort) changed successfully
```

The data types in most of the columns could not be converted because of the presence of .., so we will replace them with NaN and then handle the missing values

In [8]:
```
try:
    df.replace("..", np.nan, inplace=True)
    print("Replacement successful")
except Exception as e:
    print(e)
```

```
Replacement successful
```

In [9]:
```
for i in df.columns[3:]: #the first three columns have the correct Data type
    try:
        df[i] = df[i].astype("float64")
        print(f"Data type for {i} changed successfully")
    except Exception as e:
        print(e)
```

```
Data type for Current account balance (% of GDP) changed successfully
Data type for Exports of goods and services (% of GDP) changed successfully
Data type for Foreign direct investment, net inflows (% of GDP) changed successfully
Data type for Foreign direct investment, net outflows (% of GDP) changed successfully
Data type for GDP (current US$) changed successfully
Data type for GDP growth (annual %) changed successfully
Data type for GDP per capita (current US$) changed successfully
Data type for GDP per capita growth (annual %) changed successfully
Data type for General government final consumption expenditure (% of GDP) changed successfully
Data type for Government expenditure on education, total (% of GDP) changed successfully
Data type for Government expenditure per student, primary (% of GDP per capita) changed successfully
Data type for Government expenditure per student, secondary (% of GDP per capita) changed successfully
Data type for Government expenditure per student, tertiary (% of GDP per capita) changed successfully
Data type for Gross savings (% of GDP) changed successfully
Data type for Imports of goods and services (% of GDP) changed successfully
Data type for GNI (current US$) changed successfully
Data type for GNI growth (annual %) changed successfully
Data type for Employment to population ratio, 15+, total (%) (national estimate) changed successfully
Data type for Employment to population ratio, 15+, female (%) (national estimate) changed successfully
Data type for Employment to population ratio, 15+, male (%) (national estimate) changed successfully
Data type for Unemployment, total (% of total labor force) (national estimate) changed successfully
Data type for Employment in services (% of total employment) (modeled ILO estimate) changed successfully
Data type for Employment in industry (% of total employment) (modeled ILO estimate) changed successfully
Data type for Employment in agriculture (% of total employment) (modeled ILO estimate) changed successfully
Data type for Poverty headcount ratio at national poverty lines (% of population) changed successfully
Data type for Multidimensional poverty headcount ratio (UNDP) (% of population) changed successfully
Data type for Fixed broadband subscriptions changed successfully
Data type for Fixed telephone subscriptions changed successfully
Data type for Mobile cellular subscriptions changed successfully
Data type for Secure Internet servers changed successfully
Data type for High-technology exports (current US$) changed successfully
Data type for High-technology exports (% of manufactured exports) changed successfully
Data type for Research and development expenditure (% of GDP) changed successfully
Data type for Scientific and technical journal articles changed successfully
Data type for Air transport, freight (million ton-km) changed successfully
Data type for Air transport, passengers carried changed successfully
Data type for Rail lines (total route-km) changed successfully
Data type for Railways, goods transported (million ton-km) changed successfully
Data type for Railways, passengers carried (million passenger-km) changed successfully
Data type for Community health workers (per 1,000 people) changed successfully
Data type for Current health expenditure (% of GDP) changed successfully
Data type for Current health expenditure per capita (current US$) changed successfully
Data type for Hospital beds (per 1,000 people) changed successfully
Data type for Physicians (per 1,000 people) changed successfully
Data type for Life expectancy at birth, total (years) changed successfully
Data type for Probability of dying among adolescents ages 10-14 years (per 1,000) changed successfully
Data type for Probability of dying among adolescents ages 15-19 years (per 1,000) changed successfully
Data type for Probability of dying among children ages 5-9 years (per 1,000) changed successfully
Data type for Probability of dying among youth ages 20-24 years (per 1,000) changed successfully
Data type for Survival to age 65, female (% of cohort) changed successfully
Data type for Survival to age 65, male (% of cohort) changed successfully
```

In [10]:
```
(df.isnull().sum()/len(df))*100
```

Out[10]:
```
Country Name                                                      0.000000
Country Code                                                      0.000000
Time                                                              0.000000
Current account balance (% of GDP)                               10.230179
Exports of goods and services (% of GDP)                          2.493606
Foreign direct investment, net inflows (% of GDP)                 2.941176
Foreign direct investment, net outflows (% of GDP)                7.544757
GDP (current US$)                                                 0.191816
GDP growth (annual %)                                             0.063939
GDP per capita (current US$)                                      0.191816
GDP per capita growth (annual %)                                  0.063939
General government final consumption expenditure (% of GDP)       3.452685
Government expenditure on education, total (% of GDP)            18.286445
Government expenditure per student, primary (% of GDP per capita)  57.736573
Government expenditure per student, secondary (% of GDP per capita)  60.549872
Government expenditure per student, tertiary (% of GDP per capita)  59.974425
Gross savings (% of GDP)                                         10.805627
Imports of goods and services (% of GDP)                          2.493606
GNI (current US$)                                                 1.662404
GNI growth (annual %)                                            21.483376
Employment to population ratio, 15+, total (%) (national estimate)  21.803069
Employment to population ratio, 15+, female (%) (national estimate)  23.657289
Employment to population ratio, 15+, male (%) (national estimate)  23.849105
Unemployment, total (% of total labor force) (national estimate)  17.583120
Employment in services (% of total employment) (modeled ILO estimate)  1.470588
Employment in industry (% of total employment) (modeled ILO estimate)  1.470588
Employment in agriculture (% of total employment) (modeled ILO estimate)  1.470588
Poverty headcount ratio at national poverty lines (% of population)  74.488491
Multidimensional poverty headcount ratio (UNDP) (% of population)  98.145780
Fixed broadband subscriptions                                     7.672634
Fixed telephone subscriptions                                     0.639386
Mobile cellular subscriptions                                     0.191816
Secure Internet servers                                          44.437340
High-technology exports (current US$)                            37.276215
High-technology exports (% of manufactured exports)              37.276215
Research and development expenditure (% of GDP)                  34.271100
Scientific and technical journal articles                        10.038363
Air transport, freight (million ton-km)                          11.381074
Air transport, passengers carried                                10.613811
Rail lines (total route-km)                                      48.976982
Railways, goods transported (million ton-km)                     43.414322
Railways, passengers carried (million passenger-km)              42.327366
Community health workers (per 1,000 people)                     96.035806
Current health expenditure (% of GDP)                             6.905371
Current health expenditure per capita (current US$)               6.969309
Hospital beds (per 1,000 people)                                 25.191816
Physicians (per 1,000 people)                                    34.079284
Life expectancy at birth, total (years)                           0.000000
Probability of dying among adolescents ages 10-14 years (per 1,000)  2.941176
Probability of dying among adolescents ages 15-19 years (per 1,000)  2.941176
Probability of dying among children ages 5-9 years (per 1,000)    2.941176
Probability of dying among youth ages 20-24 years (per 1,000)     2.941176
Survival to age 65, female (% of cohort)                          0.000000
Survival to age 65, male (% of cohort)                            0.000000
dtype: float64
```

There are a lot of null values in the dataset, so in order to deal with them I have a set of guidelines:-

i) if the null values are above 50%, we will drop the column

ii) if the null values are between 10% and 50%, we will impute the values

iii) if the null values are less that 10%, we will drop just the null rows

In [11]:
```
null_values = list((df.isnull().sum()/len(df))*100)

cols_to_delete = []

for i in range(len(null_values)):
    col = df.columns[i]
    if null_values[i] >= 50:
        cols_to_delete.append(col)
        print(f"{i}: {col} marked for deletion")
    elif 10 <= null_values[i] < 50:
        df[col] = df[col].fillna(df[col].median())
        print(f"Values in column '{col}' imputed with median")
```

```
Values in column 'Current account balance (% of GDP)' imputed with median
Values in column 'Government expenditure on education, total (% of GDP)' imputed with median
13: Government expenditure per student, primary (% of GDP per capita) marked for deletion
14: Government expenditure per student, secondary (% of GDP per capita) marked for deletion
15: Government expenditure per student, tertiary (% of GDP per capita) marked for deletion
Values in column 'Gross savings (% of GDP)' imputed with median
Values in column 'GNI growth (annual %)' imputed with median
Values in column 'Employment to population ratio, 15+, total (%) (national estimate)' imputed with median
Values in column 'Employment to population ratio, 15+, female (%) (national estimate)' imputed with median
Values in column 'Employment to population ratio, 15+, male (%) (national estimate)' imputed with median
Values in column 'Unemployment, total (% of total labor force) (national estimate)' imputed with median
27: Poverty headcount ratio at national poverty lines (% of population) marked for deletion
28: Multidimensional poverty headcount ratio (UNDP) (% of population) marked for deletion
Values in column 'Secure Internet servers' imputed with median
Values in column 'High-technology exports (current US$)' imputed with median
Values in column 'High-technology exports (% of manufactured exports)' imputed with median
Values in column 'Research and development expenditure (% of GDP)' imputed with median
Values in column 'Scientific and technical journal articles' imputed with median
Values in column 'Air transport, freight (million ton-km)' imputed with median
Values in column 'Air transport, passengers carried' imputed with median
Values in column 'Rail lines (total route-km)' imputed with median
Values in column 'Railways, goods transported (million ton-km)' imputed with median
Values in column 'Railways, passengers carried (million passenger-km)' imputed with median
42: Community health workers (per 1,000 people) marked for deletion
Values in column 'Hospital beds (per 1,000 people)' imputed with median
Values in column 'Physicians (per 1,000 people)' imputed with median
```

In [12]:
```
for i in cols_to_delete:
    df.drop(i, axis=1, inplace=True)
    print(f"Column {i} deleted successfully")
```

```
Column Government expenditure per student, primary (% of GDP per capita) deleted successfully
Column Government expenditure per student, secondary (% of GDP per capita) deleted successfully
Column Government expenditure per student, tertiary (% of GDP per capita) deleted successfully
Column Poverty headcount ratio at national poverty lines (% of population) deleted successfully
Column Multidimensional poverty headcount ratio (UNDP) (% of population) deleted successfully
Column Community health workers (per 1,000 people) deleted successfully
```

Now we will remove all the null values

In [13]:
```
df.dropna(inplace=True)
```

In [14]:
```
(df.isnull().sum()/len(df))*100
```

Out[14]:
```
Country Name                                                        0.0
Country Code                                                        0.0
Time                                                                0.0
Current account balance (% of GDP)                                  0.0
Exports of goods and services (% of GDP)                            0.0
Foreign direct investment, net inflows (% of GDP)                   0.0
Foreign direct investment, net outflows (% of GDP)                  0.0
GDP (current US$)                                                   0.0
GDP growth (annual %)                                               0.0
GDP per capita (current US$)                                        0.0
GDP per capita growth (annual %)                                    0.0
General government final consumption expenditure (% of GDP)         0.0
Government expenditure on education, total (% of GDP)               0.0
Gross savings (% of GDP)                                            0.0
Imports of goods and services (% of GDP)                            0.0
GNI (current US$)                                                   0.0
GNI growth (annual %)                                               0.0
Employment to population ratio, 15+, total (%) (national estimate)  0.0
Employment to population ratio, 15+, female (%) (national estimate) 0.0
Employment to population ratio, 15+, male (%) (national estimate)   0.0
Unemployment, total (% of total labor force) (national estimate)    0.0
Employment in services (% of total employment) (modeled ILO estimate)  0.0
Employment in industry (% of total employment) (modeled ILO estimate)  0.0
Employment in agriculture (% of total employment) (modeled ILO estimate)  0.0
Fixed broadband subscriptions                                       0.0
Fixed telephone subscriptions                                       0.0
Mobile cellular subscriptions                                       0.0
Secure Internet servers                                             0.0
High-technology exports (current US$)                               0.0
High-technology exports (% of manufactured exports)                 0.0
Research and development expenditure (% of GDP)                     0.0
Scientific and technical journal articles                           0.0
Air transport, freight (million ton-km)                             0.0
Air transport, passengers carried                                   0.0
Rail lines (total route-km)                                         0.0
Railways, goods transported (million ton-km)                        0.0
Railways, passengers carried (million passenger-km)                 0.0
Current health expenditure (% of GDP)                               0.0
Current health expenditure per capita (current US$)                 0.0
Hospital beds (per 1,000 people)                                    0.0
Physicians (per 1,000 people)                                       0.0
Life expectancy at birth, total (years)                             0.0
Probability of dying among adolescents ages 10-14 years (per 1,000) 0.0
Probability of dying among adolescents ages 15-19 years (per 1,000) 0.0
Probability of dying among children ages 5-9 years (per 1,000)      0.0
Probability of dying among youth ages 20-24 years (per 1,000)       0.0
Survival to age 65, female (% of cohort)                            0.0
Survival to age 65, male (% of cohort)                              0.0
dtype: float64
```

There are no null values left in the dataset

In [15]:
```
df.shape
```

Out[15]:
```
(1245, 48)
```

We approximately deleted 300 rows and 7 columns from our dataset. Now we will handle the outliers

In [16]:
```
df1 = df
for i in df.columns[3:]:
    q1 = df[i].quantile(.25)
    q3 = df[i].quantile(.75)
    iqr = q3 - q1
    lower = q1 - 1.5*iqr
    upper = q3 + 1.5*iqr
    df1 = df[(df[i] >= lower) & (df[i] <= upper)]
    print(f"Outliers for {i} deleted successfully")
```

```
Outliers for Current account balance (% of GDP) deleted successfully
Outliers for Exports of goods and services (% of GDP) deleted successfully
Outliers for Foreign direct investment, net inflows (% of GDP) deleted successfully
Outliers for Foreign direct investment, net outflows (% of GDP) deleted successfully
Outliers for GDP (current US$) deleted successfully
Outliers for GDP growth (annual %) deleted successfully
Outliers for GDP per capita (current US$) deleted successfully
Outliers for GDP per capita growth (annual %) deleted successfully
Outliers for General government final consumption expenditure (% of GDP) deleted successfully
Outliers for Government expenditure on education, total (% of GDP) deleted successfully
Outliers for Gross savings (% of GDP) deleted successfully
Outliers for Imports of goods and services (% of GDP) deleted successfully
Outliers for GNI (current US$) deleted successfully
Outliers for GNI growth (annual %) deleted successfully
Outliers for Employment to population ratio, 15+, total (%) (national estimate) deleted successfully
Outliers for Employment to population ratio, 15+, female (%) (national estimate) deleted successfully
Outliers for Employment to population ratio, 15+, male (%) (national estimate) deleted successfully
Outliers for Unemployment, total (% of total labor force) (national estimate) deleted successfully
Outliers for Employment in services (% of total employment) (modeled ILO estimate) deleted successfully
Outliers for Employment in industry (% of total employment) (modeled ILO estimate) deleted successfully
Outliers for Employment in agriculture (% of total employment) (modeled ILO estimate) deleted successfully
Outliers for Fixed broadband subscriptions deleted successfully
Outliers for Fixed telephone subscriptions deleted successfully
Outliers for Mobile cellular subscriptions deleted successfully
Outliers for Secure Internet servers deleted successfully
Outliers for High-technology exports (current US$) deleted successfully
Outliers for High-technology exports (% of manufactured exports) deleted successfully
Outliers for Research and development expenditure (% of GDP) deleted successfully
Outliers for Scientific and technical journal articles deleted successfully
Outliers for Air transport, freight (million ton-km) deleted successfully
Outliers for Air transport, passengers carried deleted successfully
Outliers for Rail lines (total route-km) deleted successfully
Outliers for Railways, goods transported (million ton-km) deleted successfully
Outliers for Railways, passengers carried (million passenger-km) deleted successfully
Outliers for Current health expenditure (% of GDP) deleted successfully
Outliers for Current health expenditure per capita (current US$) deleted successfully
Outliers for Hospital beds (per 1,000 people) deleted successfully
Outliers for Physicians (per 1,000 people) deleted successfully
Outliers for Life expectancy at birth, total (years) deleted successfully
Outliers for Probability of dying among adolescents ages 10-14 years (per 1,000) deleted successfully
Outliers for Probability of dying among adolescents ages 15-19 years (per 1,000) deleted successfully
Outliers for Probability of dying among children ages 5-9 years (per 1,000) deleted successfully
Outliers for Probability of dying among youth ages 20-24 years (per 1,000) deleted successfully
Outliers for Survival to age 65, female (% of cohort) deleted successfully
Outliers for Survival to age 65, male (% of cohort) deleted successfully
```

In [17]:
```
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1176 entries, 5 to 1562
Data columns (total 48 columns):
 #   Column                                                                Non-Null Count  Dtype
---  ------                                                                --------------  -----
 0   Country Name                                                          1176 non-null   object
 1   Country Code                                                          1176 non-null   object
 2   Time                                                                  1176 non-null   int64
 3   Current account balance (% of GDP)                                    1176 non-null   float64
 4   Exports of goods and services (% of GDP)                              1176 non-null   float64
 5   Foreign direct investment, net inflows (% of GDP)                     1176 non-null   float64
 6   Foreign direct investment, net outflows (% of GDP)                    1176 non-null   float64
 7   GDP (current US$)                                                     1176 non-null   float64
 8   GDP growth (annual %)                                                 1176 non-null   float64
 9   GDP per capita (current US$)                                          1176 non-null   float64
 10  GDP per capita growth (annual %)                                      1176 non-null   float64
 11  General government final consumption expenditure (% of GDP)           1176 non-null   float64
 12  Government expenditure on education, total (% of GDP)                 1176 non-null   float64
 13  Gross savings (% of GDP)                                              1176 non-null   float64
 14  Imports of goods and services (% of GDP)                              1176 non-null   float64
 15  GNI (current US$)                                                     1176 non-null   float64
 16  GNI growth (annual %)                                                 1176 non-null   float64
 17  Employment to population ratio, 15+, total (%) (national estimate)    1176 non-null   float64
 18  Employment to population ratio, 15+, female (%) (national estimate)   1176 non-null   float64
 19  Employment to population ratio, 15+, male (%) (national estimate)     1176 non-null   float64
 20  Unemployment, total (% of total labor force) (national estimate)      1176 non-null   float64
 21  Employment in services (% of total employment) (modeled ILO estimate) 1176 non-null   float64
 22  Employment in industry (% of total employment) (modeled ILO estimate) 1176 non-null   float64
 23  Employment in agriculture (% of total employment) (modeled ILO estimate) 1176 non-null float64
 24  Fixed broadband subscriptions                                         1176 non-null   float64
 25  Fixed telephone subscriptions                                         1176 non-null   float64
 26  Mobile cellular subscriptions                                         1176 non-null   float64
 27  Secure Internet servers                                               1176 non-null   float64
 28  High-technology exports (current US$)                                 1176 non-null   float64
 29  High-technology exports (% of manufactured exports)                   1176 non-null   float64
 30  Research and development expenditure (% of GDP)                       1176 non-null   float64
 31  Scientific and technical journal articles                             1176 non-null   float64
 32  Air transport, freight (million ton-km)                               1176 non-null   float64
 33  Air transport, passengers carried                                     1176 non-null   float64
 34  Rail lines (total route-km)                                           1176 non-null   float64
 35  Railways, goods transported (million ton-km)                          1176 non-null   float64
 36  Railways, passengers carried (million passenger-km)                   1176 non-null   float64
 37  Current health expenditure (% of GDP)                                 1176 non-null   float64
 38  Current health expenditure per capita (current US$)                   1176 non-null   float64
 39  Hospital beds (per 1,000 people)                                      1176 non-null   float64
 40  Physicians (per 1,000 people)                                         1176 non-null   float64
 41  Life expectancy at birth, total (years)                               1176 non-null   float64
 42  Probability of dying among adolescents ages 10-14 years (per 1,000)   1176 non-null   float64
 43  Probability of dying among adolescents ages 15-19 years (per 1,000)   1176 non-null   float64
 44  Probability of dying among children ages 5-9 years (per 1,000)        1176 non-null   float64
 45  Probability of dying among youth ages 20-24 years (per 1,000)         1176 non-null   float64
 46  Survival to age 65, female (% of cohort)                              1176 non-null   float64
 47  Survival to age 65, male (% of cohort)                                1176 non-null   float64
dtypes: float64(45), int64(1), object(2)
memory usage: 450.2+ KB
```

In [18]:
```
df1.describe()
```

Out[18]:

	Time	Current account balance (% of GDP)	Exports of goods and services (% of GDP)	Foreign direct investment, net inflows (% of GDP)	Foreign direct investment, net outflows (% of GDP)	GDP (current US$)	GDP growth (annual %)	GDP per capita (current US$)	GDP per capita growth (annual %)	General government final consumption expenditure (% of GDP)	...	Current health expenditure per capita (current US$)
count	1176.000000	1176.000000	1176.000000	1176.000000	1176.000000	1.176000e+03	1176.000000	1176.000000	1176.000000	1176.000000	...	1176.000000
mean	2011.312075	0.738546	42.439117	3.727269	2.791240	1.377294e+12	3.195173	25008.143169	1.898861	16.461860	...	2086.616935
std	6.201199	7.433469	30.687337	7.329715	7.312648	3.194280e+12	3.878858	22449.195583	3.736378	4.841612	...	2291.591401
min	2000.000000	-22.939294	8.221612	-40.086346	-42.285842	4.906494e+09	-17.668333	405.424902	-18.854382	5.039343	...	14.414308
25%	2006.000000	-3.125145	24.068339	1.166062	0.300421	1.467500e+11	1.490717	5268.512788	0.422649	12.286122	...	284.519798
50%	2011.000000	-0.780812	33.273210	2.367740	1.153858	3.200000e+11	3.096182	18989.981000	1.974779	16.842715	...	1114.214966
75%	2017.000000	2.952055	52.349917	4.139932	3.363189	1.150000e+12	5.298130	41019.857433	3.951721	19.958276	...	3458.926819
max	2022.000000	45.459992	228.993771	106.498789	104.591831	2.570000e+13	26.170246	108798.451200	23.304693	29.321635	...	12473.791020

8 rows × 46 columns

In [19]:
```
df = df1.copy()
```

In [20]:
```
df['3-Year GDP Growth Avg'] = df.groupby('Country Name')['GDP per capita growth (annual %)'].transform(lambda x: x.rolling(3).mea
df['5-Year Life Expectancy Avg'] = df.groupby('Country Name')['Life expectancy at birth, total (years)'].transform(lambda x: x.ro

df['Balance to GDP Ratio'] = df['Current account balance (% of GDP)'] / (df['GDP growth (annual %)'] + 1e-6)

df['Yearly GDP Growth Change'] = df.groupby('Country Name')['GDP growth (annual %)'].transform(lambda x: x.diff())
df['Yearly Life Expectancy Change'] = df.groupby('Country Name')['Life expectancy at birth, total (years)'].transform(lambda x: x

df['Cumulative GDP Growth'] = df.groupby('Country Name')['GDP growth (annual %)'].cumsum()
df['Cumulative Life Expectancy'] = df.groupby('Country Name')['Life expectancy at birth, total (years)'].cumsum()
```

In [21]:
```
from sklearn.preprocessing import MinMaxScaler

scaler = MinMaxScaler()
scaled_columns = list(df.columns[3:])
df[scaled_columns] = scaler.fit_transform(df[scaled_columns])
```

In [22]:
```
df.head()
```

Out[22]:

	Country Name	Country Code	Time	Current account balance (% of GDP)	Exports of goods and services (% of GDP)	Foreign direct investment, net inflows (% of GDP)	Foreign direct investment, net outflows (% of GDP)	GDP (current US$)	GDP growth (annual %)	GDP per capita (current US$)	...	Probability of dying among youth ages 20-24 years (per 1,000)	Survival to age 65, female (% of cohort)	Survival to age 65, male (% of cohort)	3-Year GDP Growth Avg	5-Year Life Expectanc Av
5	Pakistan	PAK	2005	0.299065	0.018255	0.283808	0.288109	0.005452	0.569017	0.003942	...	0.508671	0.325652	0.240357	NaN	Na
6	Pakistan	PAK	2006	0.274435	0.019601	0.291476	0.288357	0.006114	0.541075	0.004646	...	0.393064	0.395308	0.275636	NaN	Na
7	Pakistan	PAK	2007	0.269467	0.015416	0.294178	0.288261	0.006970	0.504422	0.005598	...	0.358382	0.413700	0.264447	0.685705	Na
8	Pakistan	PAK	2008	0.222186	0.016552	0.291815	0.288063	0.007670	0.451401	0.006293	...	0.335260	0.429449	0.261218	0.623022	Na
9	Pakistan	PAK	2009	0.304208	0.018885	0.281982	0.288156	0.007087	0.482244	0.005350	...	0.317919	0.428256	0.265656	0.590139	0.16277

5 rows × 55 columns

In [23]:
```
new_nulls = (df.isnull().sum()/len(df))*100
for i in range(len(new_nulls)):
    if new_nulls[i] > 0:
        df[df.columns[i]] = df[df.columns[i]].fillna(df[df.columns[i]].median())
        print(f"Values successfully imputed in {df.columns[i]}")
```

```
Values successfully imputed in 3-Year GDP Growth Avg
Values successfully imputed in 5-Year Life Expectancy Avg
Values successfully imputed in Yearly GDP Growth Change
Values successfully imputed in Yearly Life Expectancy Change
```

In [24]:
```
df.isnull().sum()
```

Out[24]:
```
Country Name                                                        0
Country Code                                                        0
Time                                                                0
Current account balance (% of GDP)                                  0
Exports of goods and services (% of GDP)                            0
Foreign direct investment, net inflows (% of GDP)                   0
Foreign direct investment, net outflows (% of GDP)                  0
GDP (current US$)                                                   0
GDP growth (annual %)                                               0
GDP per capita (current US$)                                        0
GDP per capita growth (annual %)                                    0
General government final consumption expenditure (% of GDP)         0
Government expenditure on education, total (% of GDP)               0
Gross savings (% of GDP)                                            0
Imports of goods and services (% of GDP)                            0
GNI (current US$)                                                   0
GNI growth (annual %)                                               0
Employment to population ratio, 15+, total (%) (national estimate)  0
Employment to population ratio, 15+, female (%) (national estimate) 0
Employment to population ratio, 15+, male (%) (national estimate)   0
Unemployment, total (% of total labor force) (national estimate)    0
Employment in services (% of total employment) (modeled ILO estimate)  0
Employment in industry (% of total employment) (modeled ILO estimate)  0
Employment in agriculture (% of total employment) (modeled ILO estimate)  0
Fixed broadband subscriptions                                       0
Fixed telephone subscriptions                                       0
Mobile cellular subscriptions                                       0
Secure Internet servers                                             0
High-technology exports (current US$)                               0
High-technology exports (% of manufactured exports)                 0
Research and development expenditure (% of GDP)                     0
Scientific and technical journal articles                           0
Air transport, freight (million ton-km)                             0
Air transport, passengers carried                                   0
Rail lines (total route-km)                                         0
Railways, goods transported (million ton-km)                        0
Railways, passengers carried (million passenger-km)                 0
Current health expenditure (% of GDP)                               0
Current health expenditure per capita (current US$)                 0
Hospital beds (per 1,000 people)                                    0
Physicians (per 1,000 people)                                       0
Life expectancy at birth, total (years)                             0
Probability of dying among adolescents ages 10-14 years (per 1,000)  0
Probability of dying among adolescents ages 15-19 years (per 1,000)  0
Probability of dying among children ages 5-9 years (per 1,000)      0
Probability of dying among youth ages 20-24 years (per 1,000)       0
Survival to age 65, female (% of cohort)                            0
Survival to age 65, male (% of cohort)                              0
3-Year GDP Growth Avg                                               0
5-Year Life Expectancy Avg                                          0
Balance to GDP Ratio                                                0
Yearly GDP Growth Change                                            0
Yearly Life Expectancy Change                                       0
Cumulative GDP Growth                                               0
Cumulative Life Expectancy                                          0
dtype: int64
```

In [25]:
```
aggregates = {}
for i in df.columns[3:]:
    if "Cumulative" not in i:
        aggregates[i] = "mean"
    else:
        aggregates[i] = "max"


aggregate_rename = {}
for i in df.columns[3:-7]:
    aggregate_rename[i] = "Average " + i
```

In [26]:
```
country_aggregates = df.groupby('Country Name').agg(aggregates).reset_index()
country_aggregates.rename(columns=aggregate_rename, inplace=True)
country_aggregates.head()
```

Out[26]:

	Country Name	Average Current account balance (% of GDP)	Average Exports of goods and services (% of GDP)	Average Foreign direct investment, net inflows (% of GDP)	Average Foreign direct investment, net outflows (% of GDP)	Average GDP (current US$)	Average GDP growth (annual %)	Average GDP per capita (current US$)	Average GDP per capita growth (annual %)	Average General government final consumption expenditure (% of GDP)	...	Average Probability of dying among youth ages 20-24 years (per 1,000)	Average Survival to age 65, female (% of cohort)	Average Survival to age 65, male (% of cohort)	3-Year GDP Growth Avg
0	Algeria	0.356630	0.104255	0.279670	0.288462	0.006479	0.466302	0.038012	0.470538	0.493519	...	0.132035	0.731982	0.694587	0.584894
1	Argentina	0.340123	0.045685	0.286121	0.289983	0.015185	0.447864	0.082597	0.469672	0.411391	...	0.204789	0.770789	0.603105	0.598248
2	Australia	0.286841	0.057104	0.295030	0.294876	0.042701	0.463070	0.437828	0.475612	0.584183	...	0.065607	0.924465	0.887300	0.595388
3	Austria	0.359772	0.193256	0.288743	0.312446	0.014248	0.434099	0.399327	0.467117	0.602613	...	0.068050	0.906305	0.806401	0.581304
4	Bahrain	0.380535	0.333603	0.304826	0.304626	0.000825	0.491881	0.193193	0.456015	0.423977	...	0.093862	0.851781	0.848976	0.566074

5 rows × 53 columns

Now lets make a new column for economic growth index, it will depend on three factors.

GDP Growth (50%): The primary indicator for economic growth, reflecting overall economic activity.

Current Account Balance (30%): A significant measure of trade and financial stability, contributing to long-term growth potential.

Life Expectancy (20%): Reflects the health and well-being of a population, indirectly impacting economic productivity and sustainability.

The proposed weights are arbitrary and could be changed or modified according to the needs

In [27]:
```
country_aggregates['Economic Growth Index'] = (
    country_aggregates['Average GDP growth (annual %)'] * 0.5 +
    country_aggregates['Average Current account balance (% of GDP)'] * 0.3 +
    country_aggregates['Average Life expectancy at birth, total (years)'] * 0.2
)
country_aggregates.head()
```

Out[27]:

	Country Name	Average Current account balance (% of GDP)	Average Exports of goods and services (% of GDP)	Average Foreign direct investment, net inflows (% of GDP)	Average Foreign direct investment, net outflows (% of GDP)	Average GDP (current US$)	Average GDP growth (annual %)	Average GDP per capita (current US$)	Average GDP per capita growth (annual %)	Average General government final consumption expenditure (% of GDP)	...	Average Survival to age 65, female (% of cohort)	Average Survival to age 65, male (% of cohort)	3-Year GDP Growth Avg	5-Year Life Expectancy Avg
0	Algeria	0.356630	0.104255	0.279670	0.288462	0.006479	0.466302	0.038012	0.470538	0.493519	...	0.731982	0.694587	0.584894	0.624571
1	Argentina	0.340123	0.045685	0.286121	0.289983	0.015185	0.447864	0.082597	0.469672	0.411391	...	0.770789	0.603105	0.598248	0.676327
2	Australia	0.286841	0.057104	0.295030	0.294876	0.042701	0.463070	0.437828	0.475612	0.584183	...	0.924465	0.887300	0.595388	0.866431
3	Austria	0.359772	0.193256	0.288743	0.312446	0.014248	0.434099	0.399327	0.467117	0.602613	...	0.906305	0.806401	0.581304	0.827204
4	Bahrain	0.380535	0.333603	0.304826	0.304626	0.000825	0.491881	0.193193	0.456015	0.423977	...	0.851781	0.848976	0.566074	0.759658

5 rows × 54 columns

In [28]:
```
country_aggregates.describe(include="all")
```

Out[28]:

	Country Name	Average Current account balance (% of GDP)	Average Exports of goods and services (% of GDP)	Average Foreign direct investment, net inflows (% of GDP)	Average Foreign direct investment, net outflows (% of GDP)	Average GDP (current US$)	Average GDP growth (annual %)	Average GDP per capita (current US$)	Average GDP per capita growth (annual %)	Average General government final consumption expenditure (% of GDP)	...	Average Survival to age 65, female (% of cohort)	Average Survival to age 65, male (% of cohort)	3-Year GDP Growth Avg
count	60	60.000000	60.000000	60.000000	60.000000	60.000000	60.000000	60.000000	60.000000	60.000000	...	60.000000	60.000000	60.000000
unique	60	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN
top	Algeria	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN
freq	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN
mean	NaN	0.342301	0.148488	0.297849	0.305253	0.048025	0.477986	0.206644	0.492274	0.452508	...	0.750988	0.620929	0.617703
std	NaN	0.087457	0.132396	0.026591	0.022609	0.111780	0.043821	0.193925	0.039610	0.195286	...	0.212838	0.248857	0.055896
min	NaN	0.229249	0.012136	0.275783	0.286403	0.000212	0.392428	0.003957	0.406319	0.019356	...	0.014448	0.004243	0.504339
25%	NaN	0.291397	0.075296	0.285269	0.291015	0.006285	0.441437	0.041417	0.470162	0.286359	...	0.717243	0.490526	0.583206
50%	NaN	0.323133	0.106933	0.291489	0.297108	0.010542	0.479973	0.110275	0.478220	0.432430	...	0.812408	0.690382	0.599209
75%	NaN	0.375067	0.193706	0.298603	0.311461	0.042685	0.501700	0.369291	0.517947	0.600167	...	0.903908	0.827537	0.653949
max	NaN	0.731011	0.843306	0.411999	0.393886	0.637725	0.601045	0.708976	0.639086	0.821999	...	0.946857	0.915748	0.831369

11 rows × 54 columns

In [29]:
```
df.describe(include="all")
```

Out[29]:

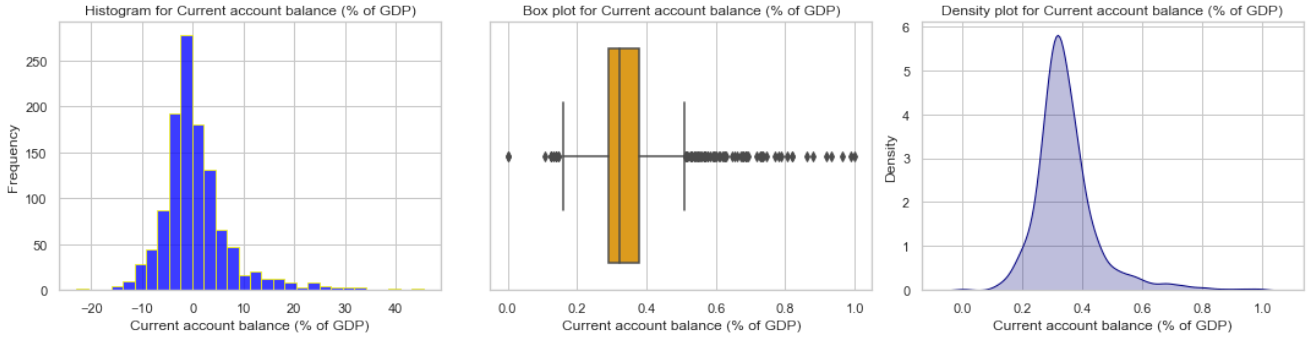
	Country Name	Country Code	Time	Current account balance (% of GDP)	Exports of goods and services (% of GDP)	Foreign direct investment, net inflows (% of GDP)	Foreign direct investment, net outflows (% of GDP)	GDP (current US$)	GDP growth (annual %)	GDP per capita (current US$)	...	Probability of dying among youth ages 20-24 years (per 1,000)	Survival t age 6 female (% of cohor
count	1176	1176	1176.000000	1176.000000	1176.000000	1176.000000	1176.000000	1176.000000	1176.000000	1176.000000	...	1176.000000	1176.00000
unique	60	60	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	Na
top	United Kingdom	GBR	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	Na
freq	23	23	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	Na
mean	NaN	NaN	2011.312075	0.346171	0.154990	0.298895	0.306902	0.053410	0.475917	0.226977	...	0.159054	0.79094
std	NaN	NaN	6.201199	0.108678	0.139000	0.050003	0.049787	0.124315	0.088480	0.207109	...	0.152201	0.16681
min	NaN	NaN	2000.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.00000
25%	NaN	NaN	2006.000000	0.289684	0.071779	0.281423	0.289944	0.005520	0.437036	0.044865	...	0.057803	0.73733
50%	NaN	NaN	2011.000000	0.323958	0.113473	0.289621	0.295754	0.012263	0.473658	0.171455	...	0.109827	0.84370
75%	NaN	NaN	2017.000000	0.378532	0.199882	0.301711	0.310796	0.044565	0.523887	0.374696	...	0.219653	0.90653
max	NaN	NaN	2022.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	...	1.000000	1.00000

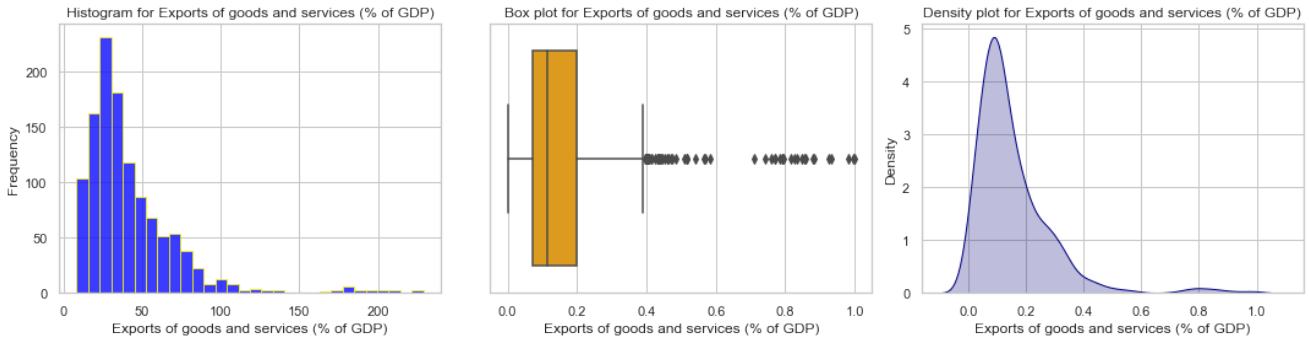
11 rows × 55 columns

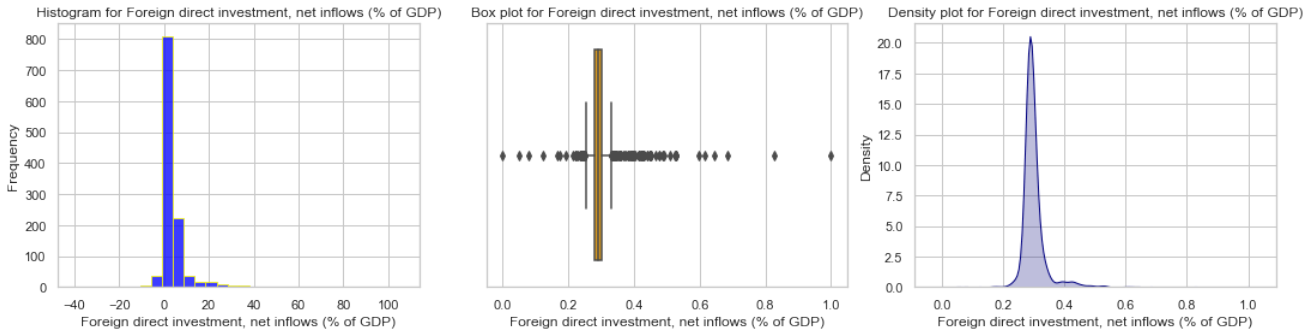
Univariate analysis

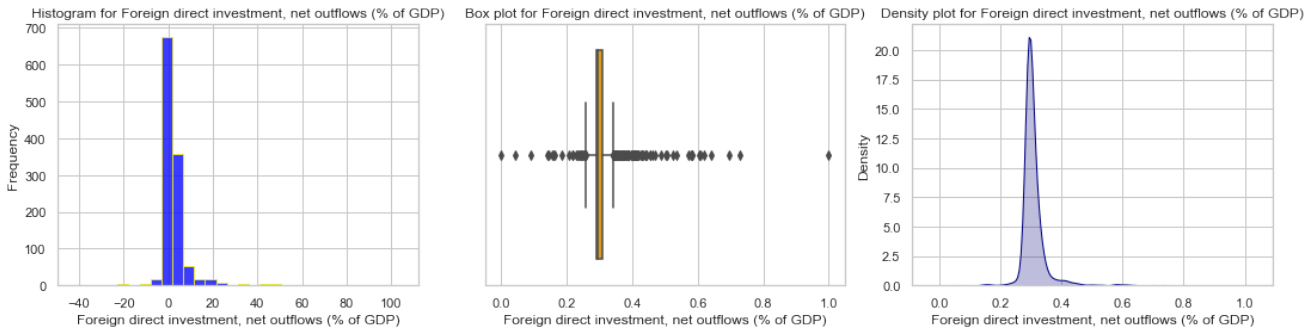
In [30]:
```
%matplotlib inline
```

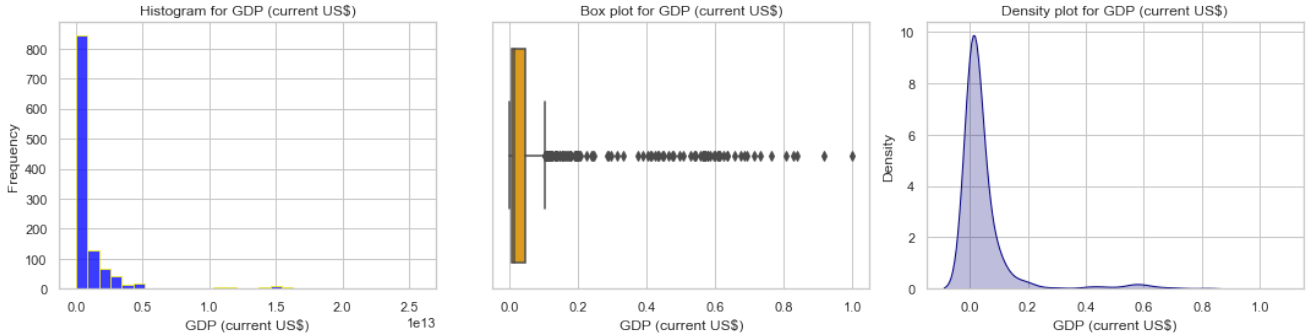
In [31]:
```
economic_metrics = list(df1.columns[3:24])
for i in economic_metrics:
    plt.figure(figsize=(15,4))

    plt.subplot(1,3,1)
    sns.histplot(data=df1[i], color="blue", edgecolor="yellow", bins=30)
    plt.xlabel(i)
    plt.ylabel("Frequency")
    plt.title(f"Histogram for {i}")

    plt.subplot(1,3,2)
    sns.boxplot(x=df[i], color="orange")
    plt.xlabel(i)
    plt.title(f"Box plot for {i}")

    plt.subplot(1,3,3)
    sns.kdeplot(data=df[i], color="navy", fill=True)
    plt.xlabel(i)
    plt.ylabel("Density")
    plt.title(f"Density plot for {i}")

    plt.tight_layout()
    plt.show()
```

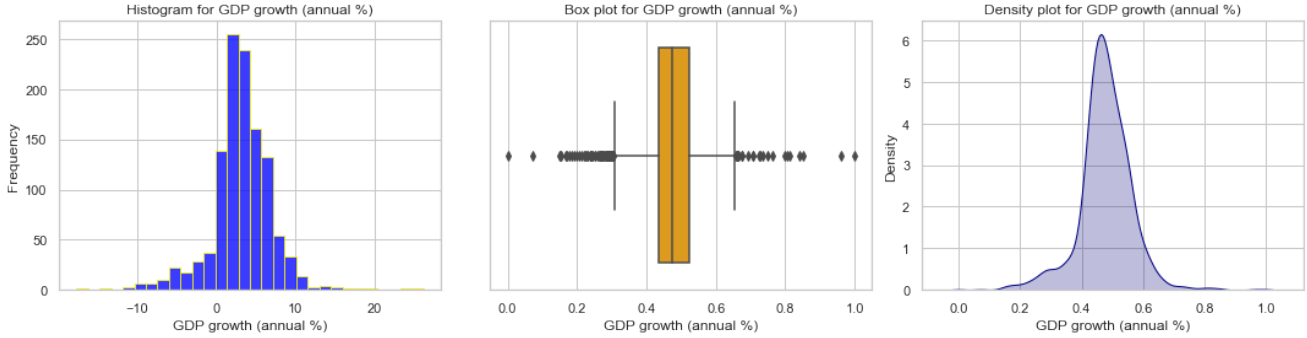


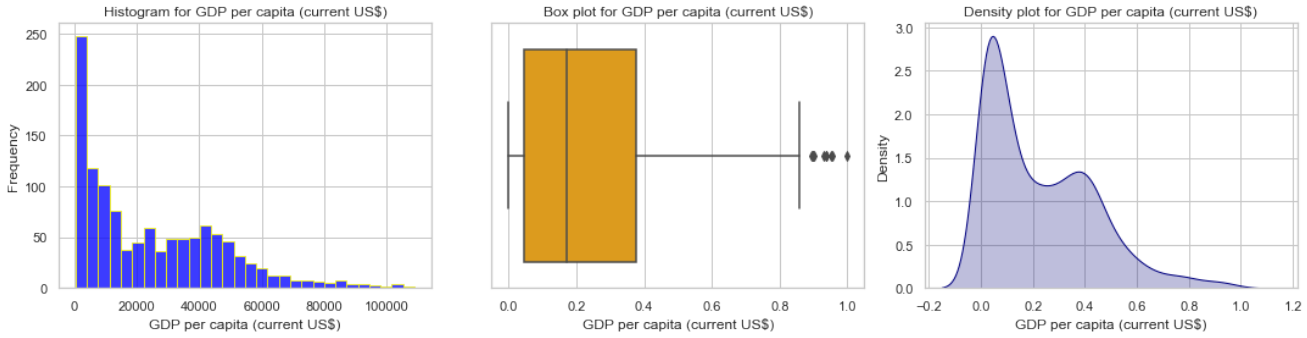


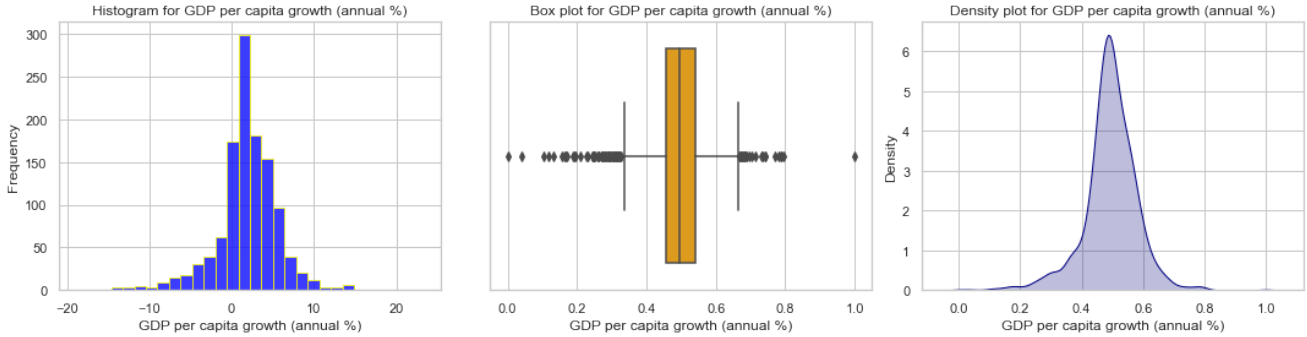


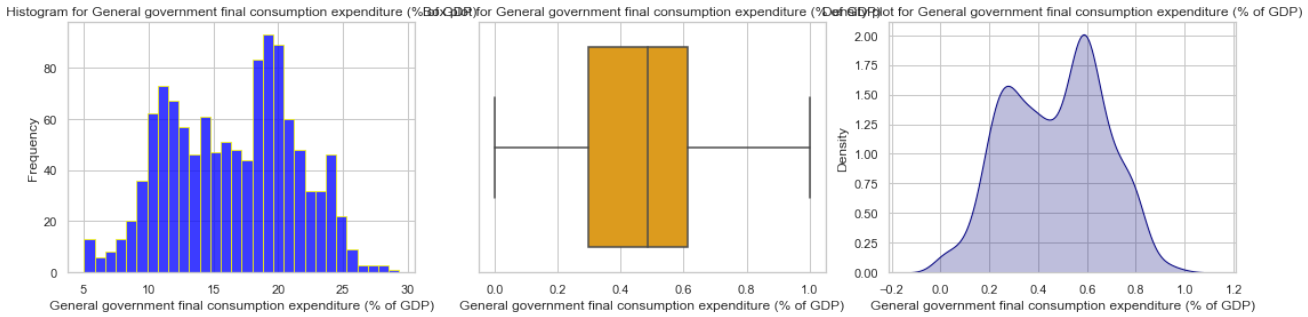

Histogram for Foreign direct investment, net outflows (% of GDP)
Frequency
Foreign direct investment, net outflows (% of GDP)
Box plot for Foreign direct investment, net outflows (% of GDP)
Foreign direct investment, net outflows (% of GDP)
Density plot for Foreign direct investment, net outflows (% of GDP)
Density
Foreign direct investment, net outflows (% of GDP)

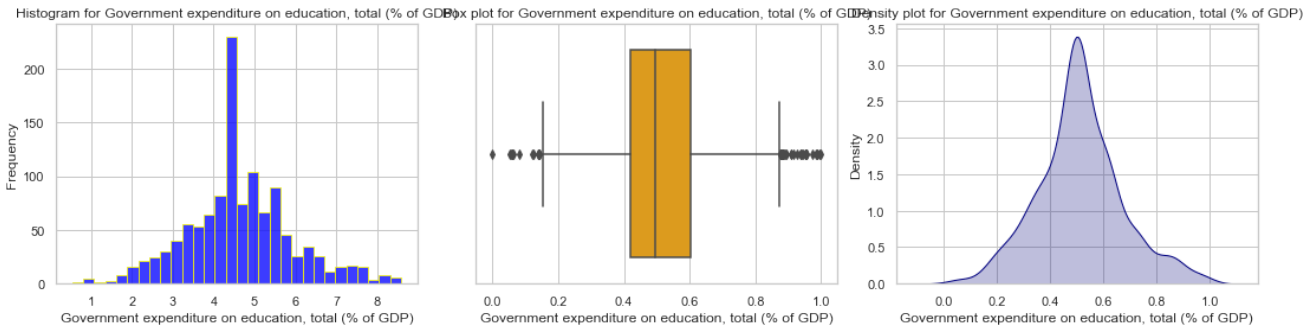

Histogram for GDP (current US$)
Frequency
GDP (current US$)
1e13
Box plot for GDP (current US$)
GDP (current US$)
Density plot for GDP (current US$)
Density
GDP (current US$)

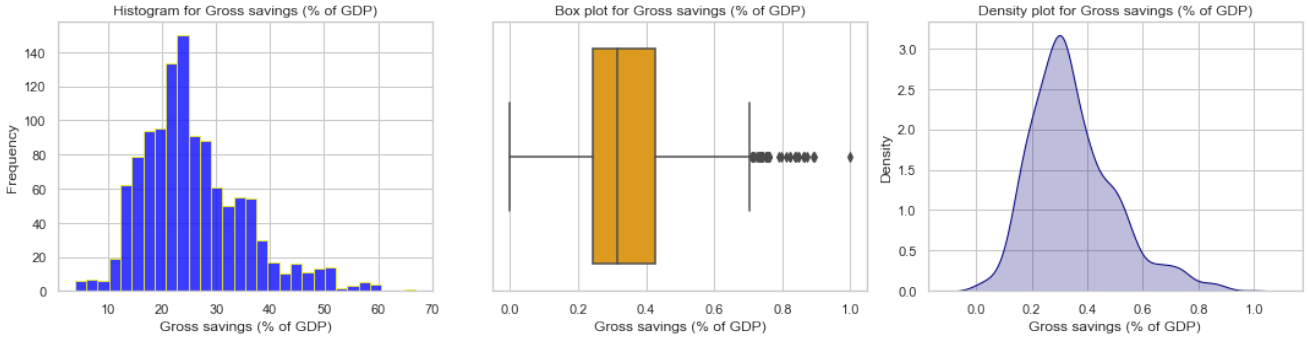

Histogram for GDP growth (annual %)
Frequency
GDP growth (annual %)
Box plot for GDP growth (annual %)
GDP growth (annual %)
Density plot for GDP growth (annual %)
Density
GDP growth (annual %)

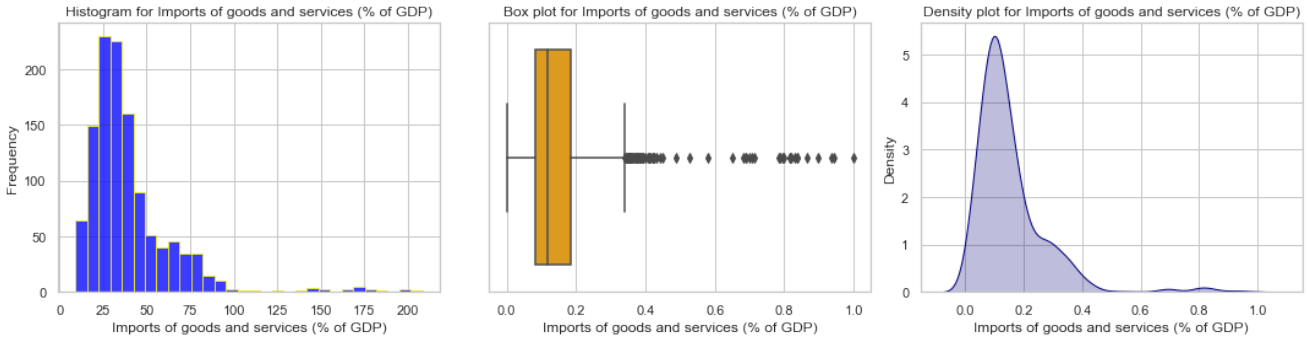

Histogram for GDP per capita (current US$)
Frequency
GDP per capita (current US$)
Box plot for GDP per capita (current US$)
GDP per capita (current US$)
Density plot for GDP per capita (current US$)
Density
GDP per capita (current US$)

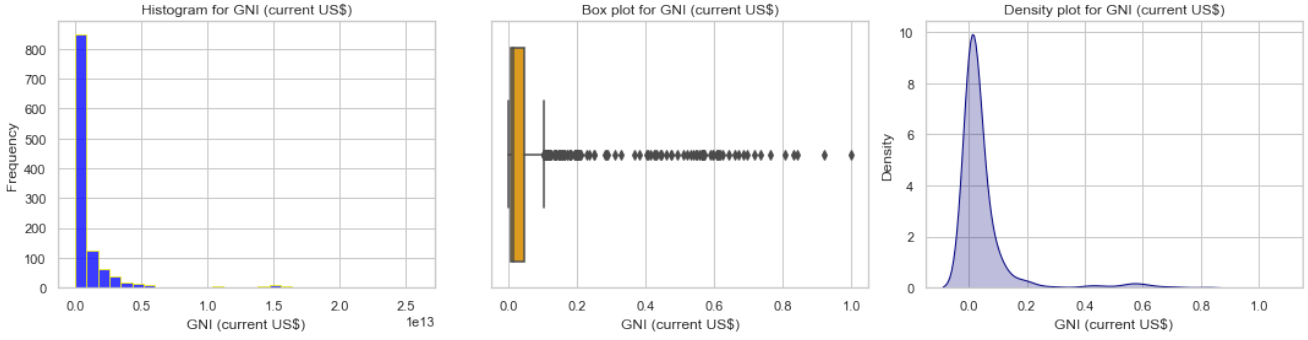

Histogram for GDP per capita growth (annual %)
Frequency
GDP per capita growth (annual %)
Box plot for GDP per capita growth (annual %)
GDP per capita growth (annual %)
Density plot for GDP per capita growth (annual %)
Density
GDP per capita growth (annual %)

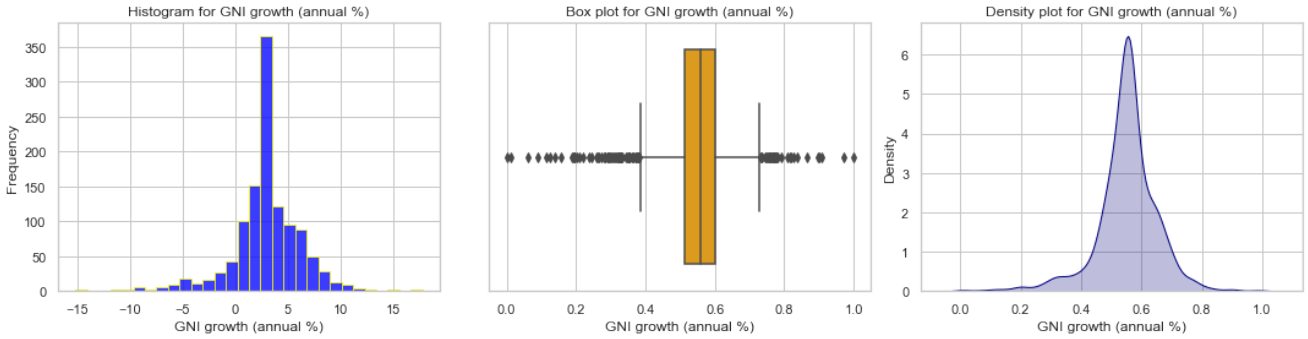

Histogram for General government final consumption expenditure (% of GDP)
Frequency
General government final consumption expenditure (% of GDP)
Box plot for General government final consumption expenditure (% of GDP)
General government final consumption expenditure (% of GDP)
Density plot for General government final consumption expenditure (% of GDP)
Density
General government final consumption expenditure (% of GDP)

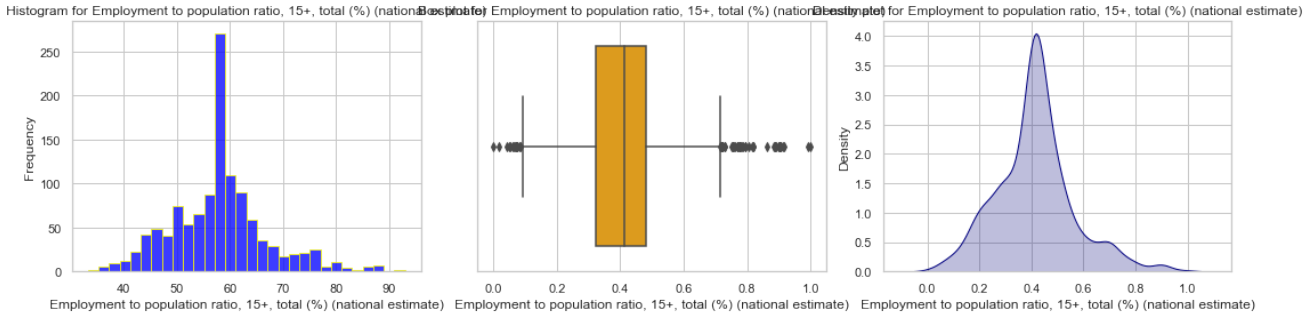

Histogram for Government expenditure on education, total (% of GDP)
Frequency
Government expenditure on education, total (% of GDP)
Box plot for Government expenditure on education, total (% of GDP)
Government expenditure on education, total (% of GDP)
Density plot for Government expenditure on education, total (% of GDP)
Density
Government expenditure on education, total (% of GDP)

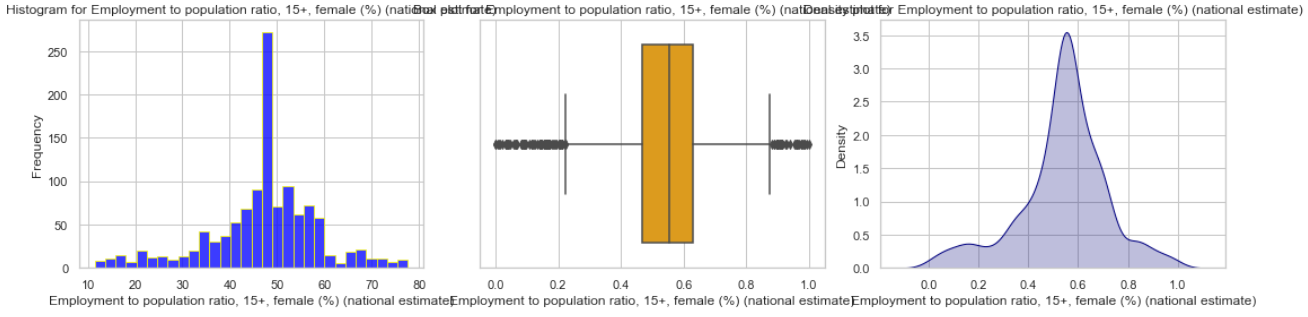

Histogram for Gross savings (% of GDP)
Frequency
Gross savings (% of GDP)
Box plot for Gross savings (% of GDP)
Gross savings (% of GDP)
Density plot for Gross savings (% of GDP)
Density
Gross savings (% of GDP)

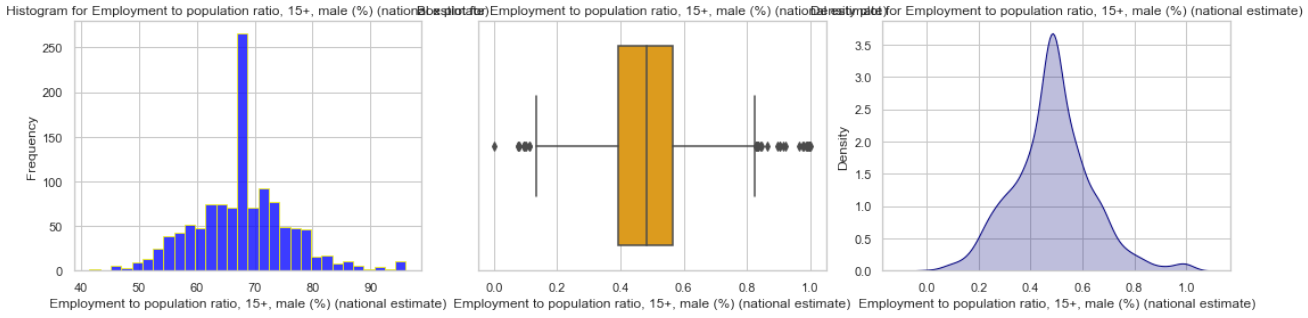

Histogram for Imports of goods and services (% of GDP)
Frequency
Imports of goods and services (% of GDP)
Box plot for Imports of goods and services (% of GDP)
Imports of goods and services (% of GDP)
Density plot for Imports of goods and services (% of GDP)
Density
Imports of goods and services (% of GDP)

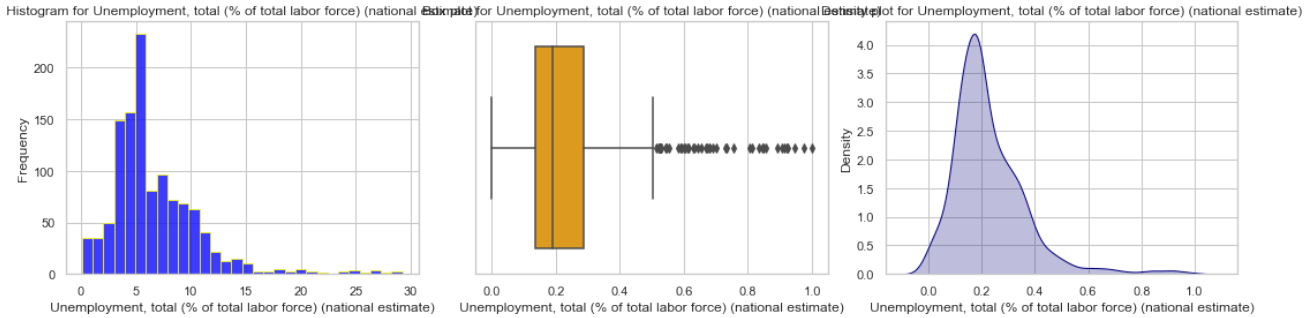

Histogram for GNI (current US$)
Frequency
GNI (current US$)
1e13
Box plot for GNI (current US$)
GNI (current US$)
Density plot for GNI (current US$)
Density
GNI (current US$)

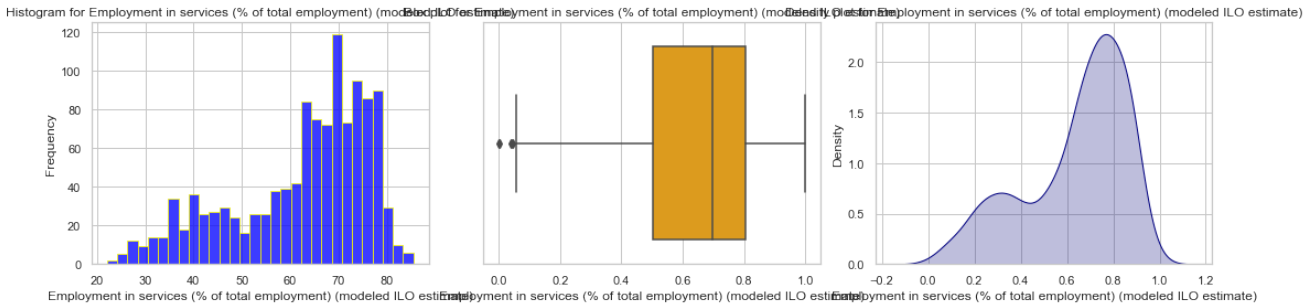

Histogram for GNI growth (annual %)
Frequency
GNI growth (annual %)
Box plot for GNI growth (annual %)
GNI growth (annual %)
Density plot for GNI growth (annual %)
Density
GNI growth (annual %)

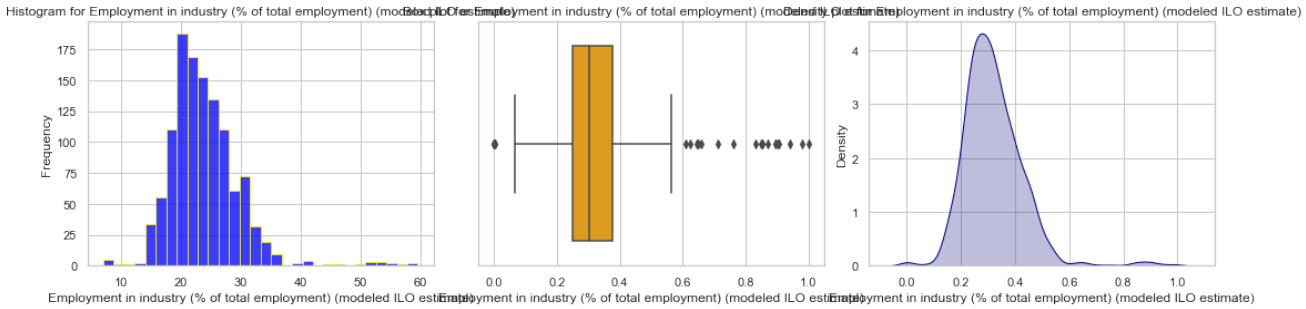

Histogram for Employment to population ratio, 15+, total (%) (national estimate)
Frequency
Employment to population ratio, 15+, total (%) (national estimate)
Box plot for Employment to population ratio, 15+, total (%) (national estimate)
Employment to population ratio, 15+, total (%) (national estimate)
Density plot for Employment to population ratio, 15+, total (%) (national estimate)
Density
Employment to population ratio, 15+, total (%) (national estimate)

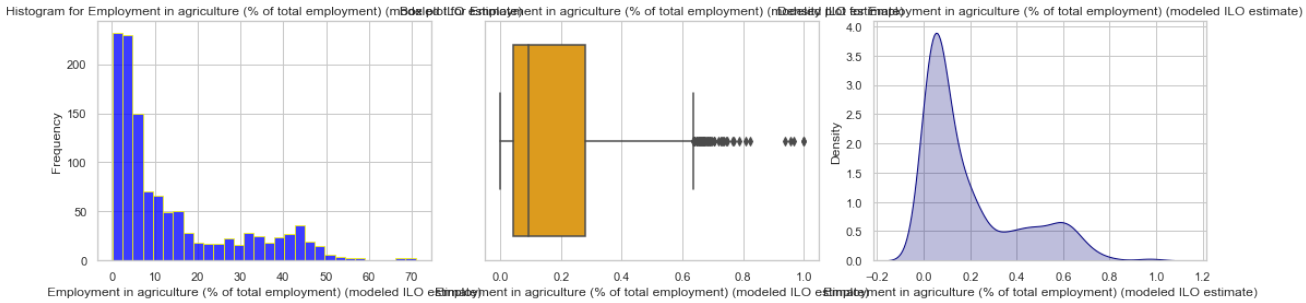

Histogram for Employment to population ratio, 15+, female (%) (national estimate)
Frequency
Employment to population ratio, 15+, female (%) (national estimate)
Box plot for Employment to population ratio, 15+, female (%) (national estimate)
Employment to population ratio, 15+, female (%) (national estimate)
Density plot for Employment to population ratio, 15+, female (%) (national estimate)
Density
Employment to population ratio, 15+, female (%) (national estimate)


Histogram for Employment to population ratio, 15+, male (%) (national estimate)
Frequency
Employment to population ratio, 15+, male (%) (national estimate)
Box plot for Employment to population ratio, 15+, male (%) (national estimate)
Employment to population ratio, 15+, male (%) (national estimate)
Density plot for Employment to population ratio, 15+, male (%) (national estimate)
Density
Employment to population ratio, 15+, male (%) (national estimate)

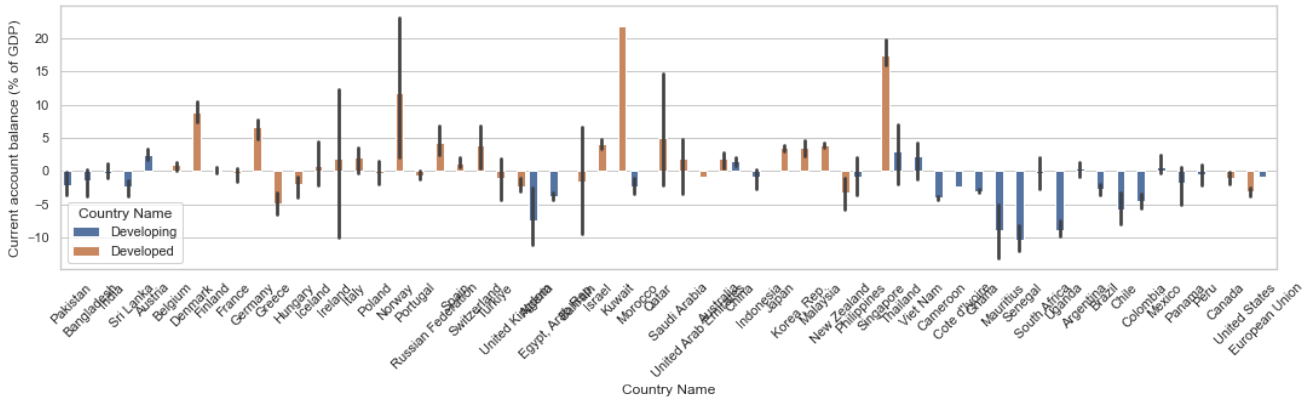

Histogram for Unemployment, total (% of total labor force) (national estimate)
Frequency
Unemployment, total (% of total labor force) (national estimate)
Box plot for Unemployment, total (% of total labor force) (national estimate)
Unemployment, total (% of total labor force) (national estimate)
Density plot for Unemployment, total (% of total labor force) (national estimate)
Density
Unemployment, total (% of total labor force) (national estimate)

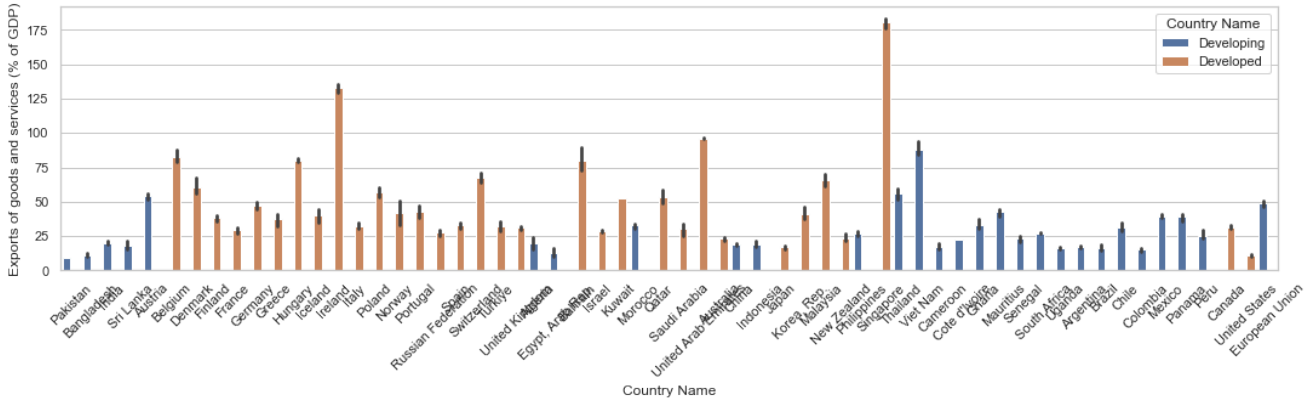

Histogram for Employment in services (% of total employment) (modeled ILO estimate)
Boxplot for Employment in services (% of total employment) (modeled ILO estimate)
Density Plot for Employment in services (% of total employment) (modeled ILO estimate)
Frequency
Density
Employment in services (% of total employment) (modeled ILO estimate)

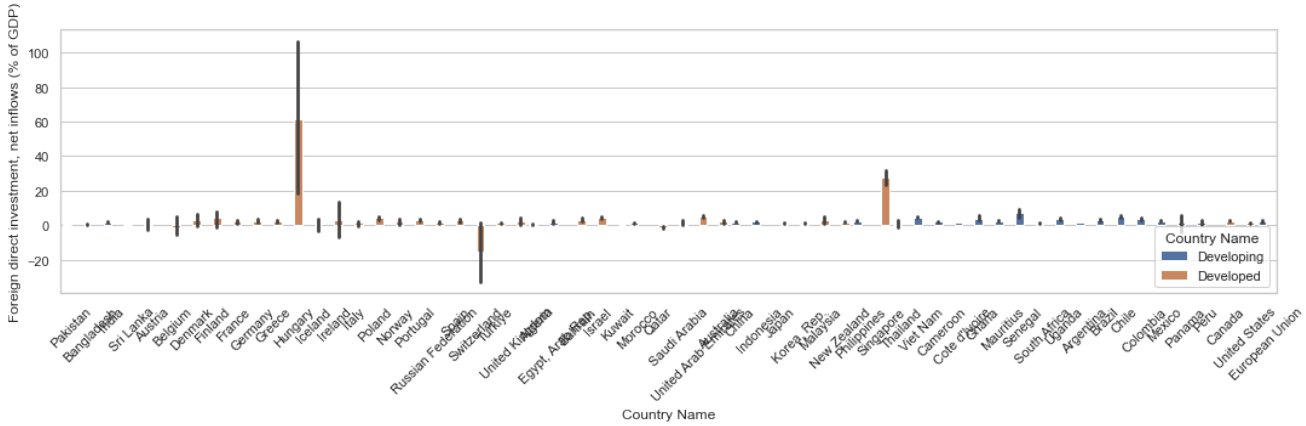

Histogram for Employment in industry (% of total employment) (modeled ILO estimate)
Boxplot for Employment in industry (% of total employment) (modeled ILO estimate)
Density Plot for Employment in industry (% of total employment) (modeled ILO estimate)
Frequency
Density
Employment in industry (% of total employment) (modeled ILO estimate)

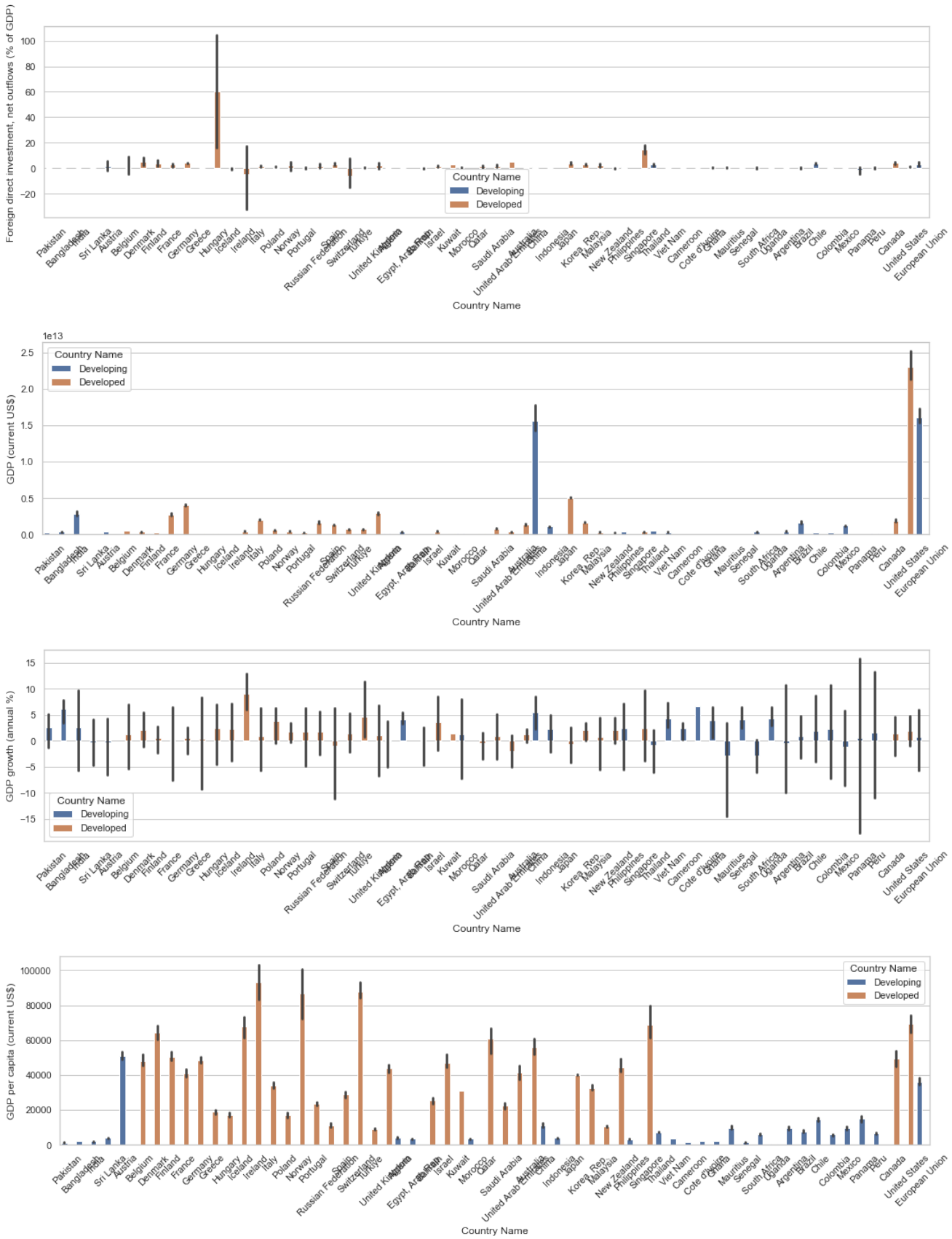

Histogram for Employment in agriculture (% of total employment) (modeled ILO estimate)
Boxplot for Employment in agriculture (% of total employment) (modeled ILO estimate)
Density Plot for Employment in agriculture (% of total employment) (modeled ILO estimate)
Frequency
Density
Employment in agriculture (% of total employment) (modeled ILO estimate)

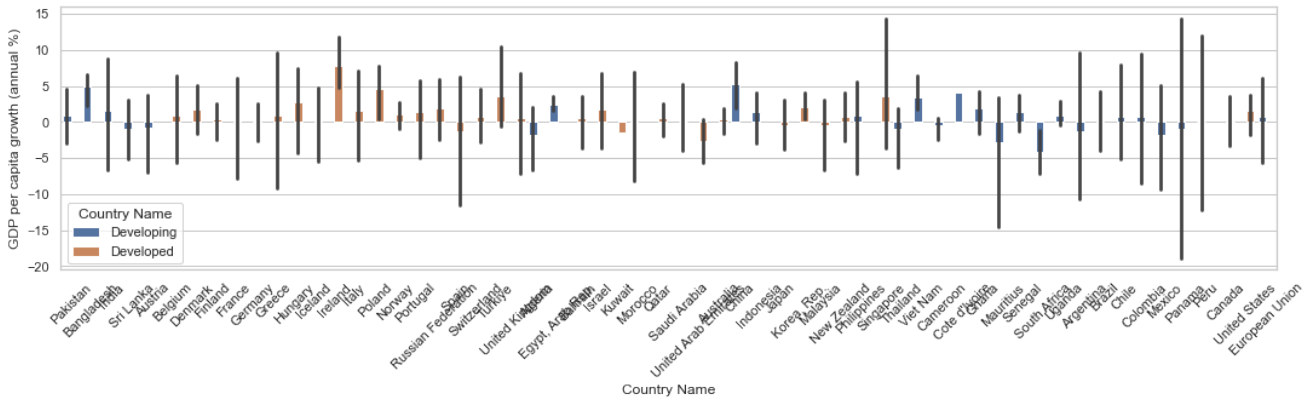
In [32]:
```
developed_countries = ['Belgium', 'Denmark', 'Finland', 'France', 'Germany', 'Greece',
       'Hungary', 'Iceland', 'Ireland', 'Italy', 'Poland', 'Norway',
       'Portugal', 'Russian Federation', 'Spain', 'Switzerland',
       'Turkiye', 'United Kingdom', 'Bahrain', 'Israel', 'Kuwait', 'Qatar', 'Saudi Arabia',
       'United Arab Emirates', 'Australia', 'Japan',
       'Korea, Rep.', 'Malaysia', 'New Zealand', 'Singapore', 'Canada', 'United States']
developing_countries = []
for i in df["Country Name"].unique():
    if i not in developed_countries:
        developing_countries.append(i)

df2 = df1[df1.Time.isin([2019,2020,2021,2022])]
for i in economic_metrics:
    plt.figure(figsize=(18,4))
    sns.barplot(data=df2, x="Country Name", y=i, hue=df2['Country Name'].apply(lambda x: 'Developed' if x in developed_countries
    plt.xticks(rotation=45)
    plt.show()
```

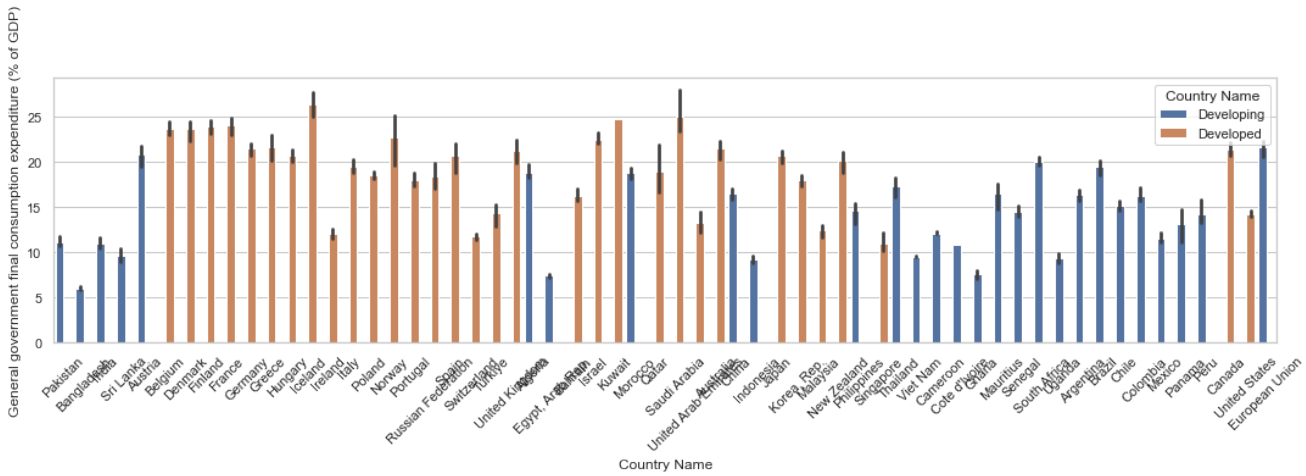


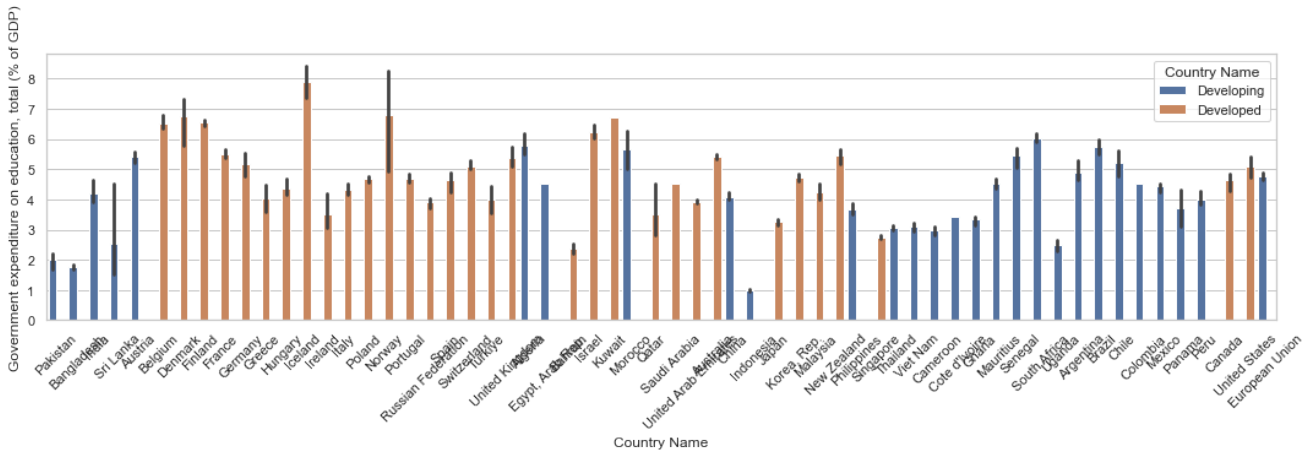


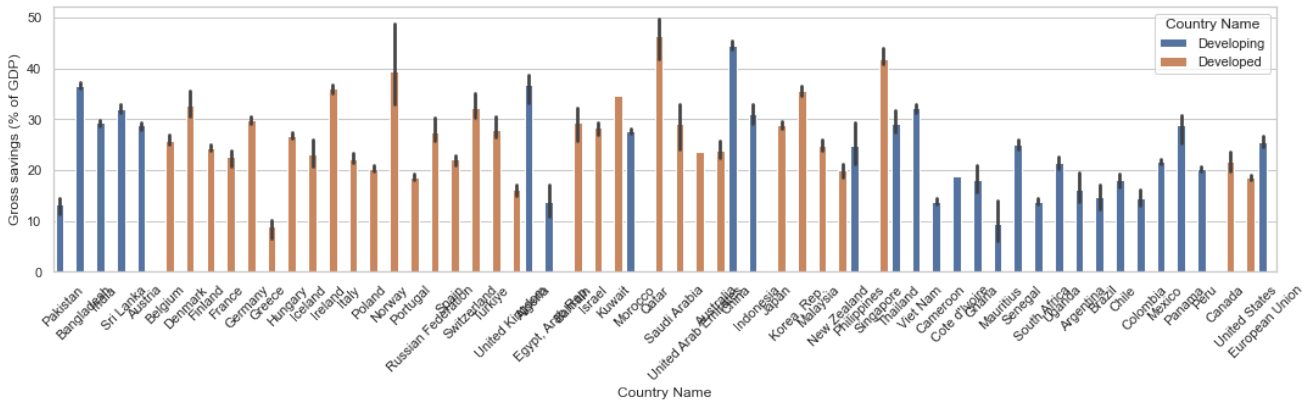


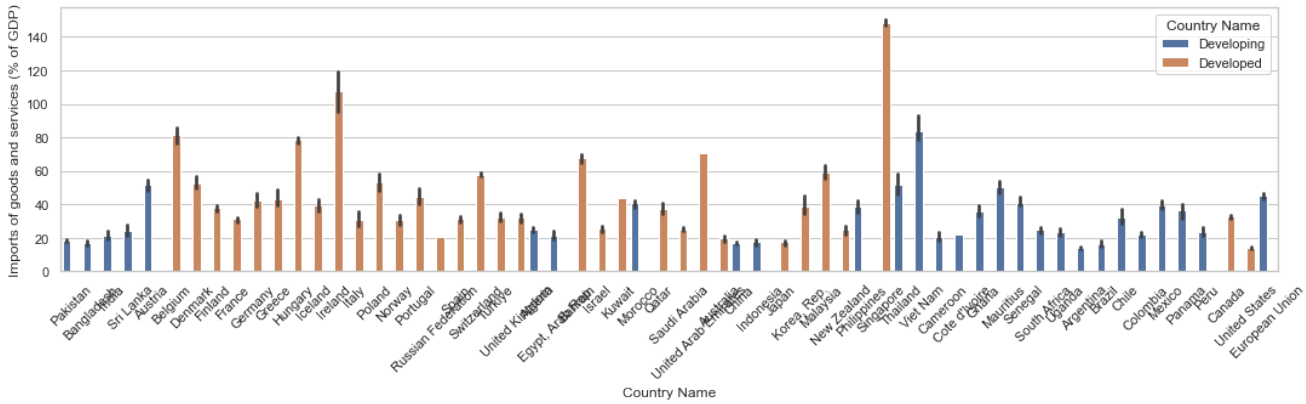

Foreign direct investment, net outflows (% of GDP)
Country Name
Developing
Developed
1e13
GDP (current US$)
GDP growth (annual %)
GDP per capita (current US$)
Pakistan
Bangladesh
India
Sri Lanka
Austria
Belgium
Denmark
Finland
France
Germany
Greece
Hungary
Iceland
Ireland
Italy
Poland
Norway
Portugal
Spain
Switzerland
Türkiye
United Kingdom
Algeria
Egypt, Arab Rep.
Iran
Israel
Kuwait
Morocco
Qatar
Saudi Arabia
Australia
United Arab Emirates
China
Indonesia
Japan
Korea, Rep.
Malaysia
New Zealand
Philippines
Singapore
Thailand
Viet Nam
Cameroon
Cote d'Ivoire
Ghana
Mauritius
Senegal
South Africa
Uganda
Argentina
Brazil
Chile
Colombia
Mexico
Panama
Peru
Canada
United States
European Union

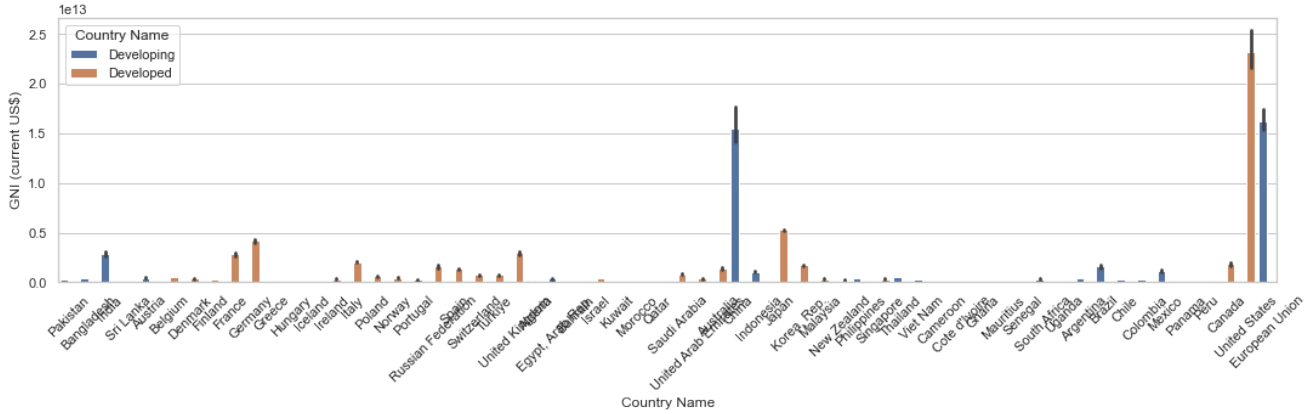

GDP per capita growth (annual %)
Country Name
Developing
Developed
Country Name

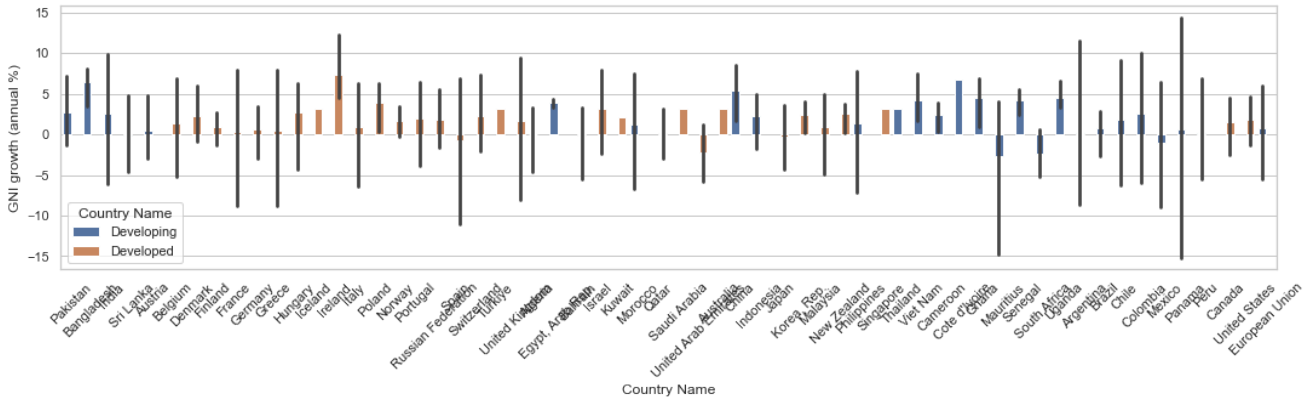

General government final consumption expenditure (% of GDP)
Country Name
Developing
Developed
Country Name

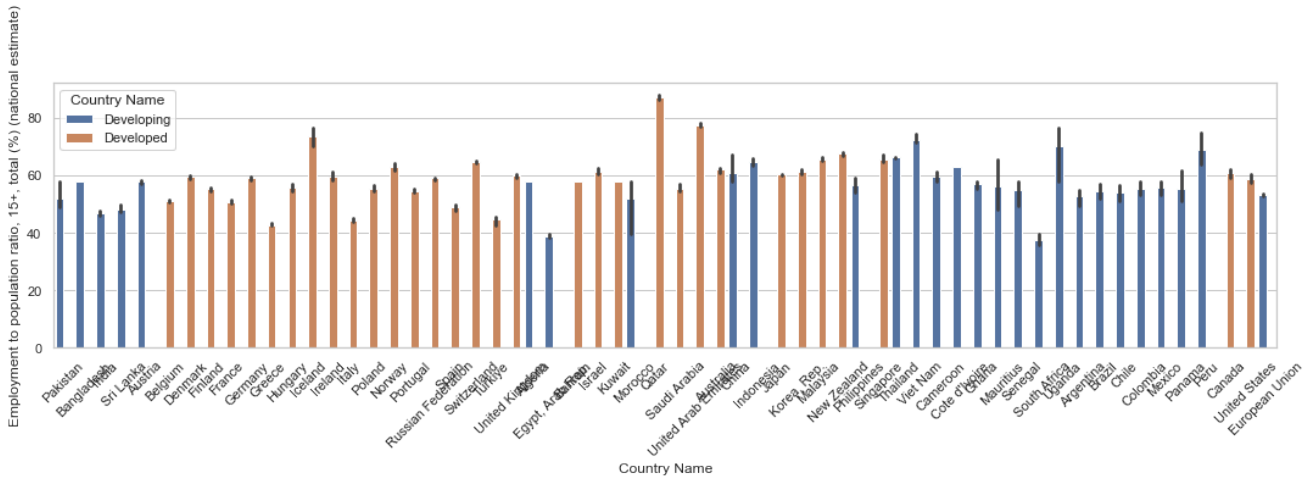

Government expenditure on education, total (% of GDP)
Country Name
Developing
Developed
Country Name

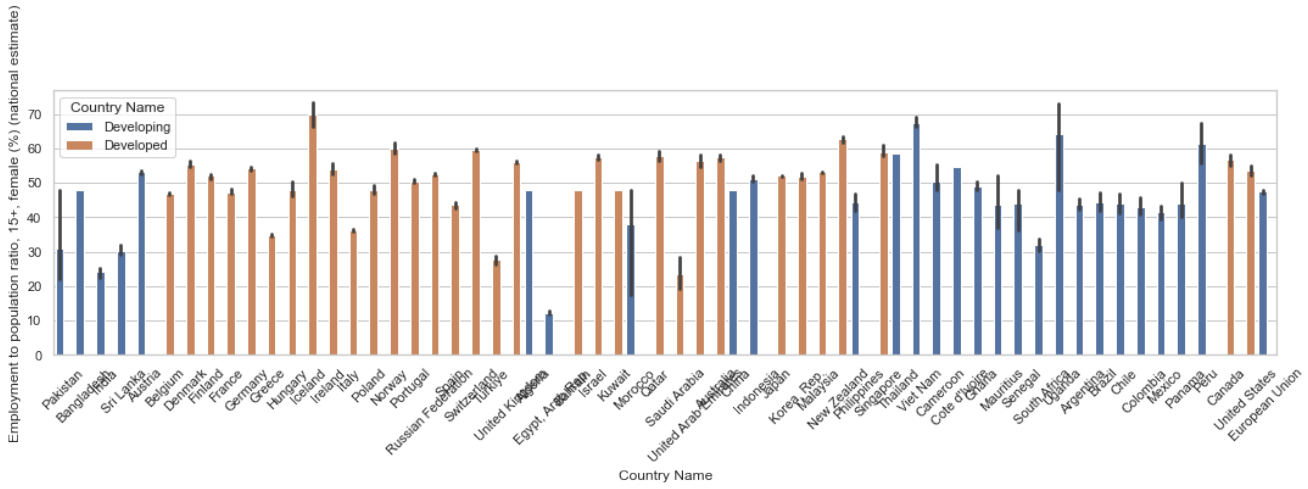

Gross savings (% of GDP)
Country Name
Developing
Developed
Country Name

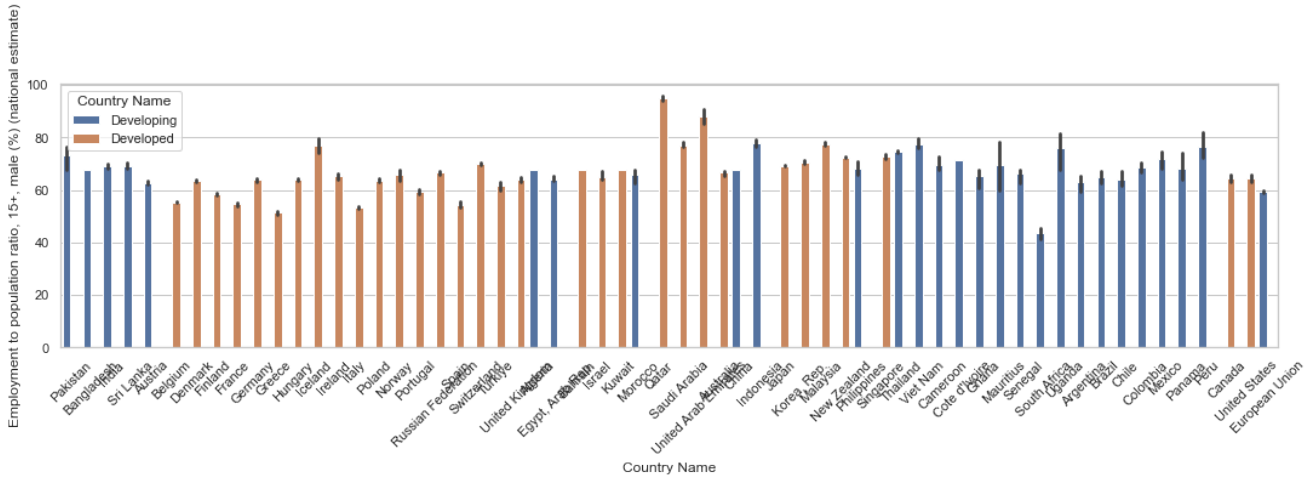

Imports of goods and services (% of GDP)
Country Name
Developing
Developed
Country Name

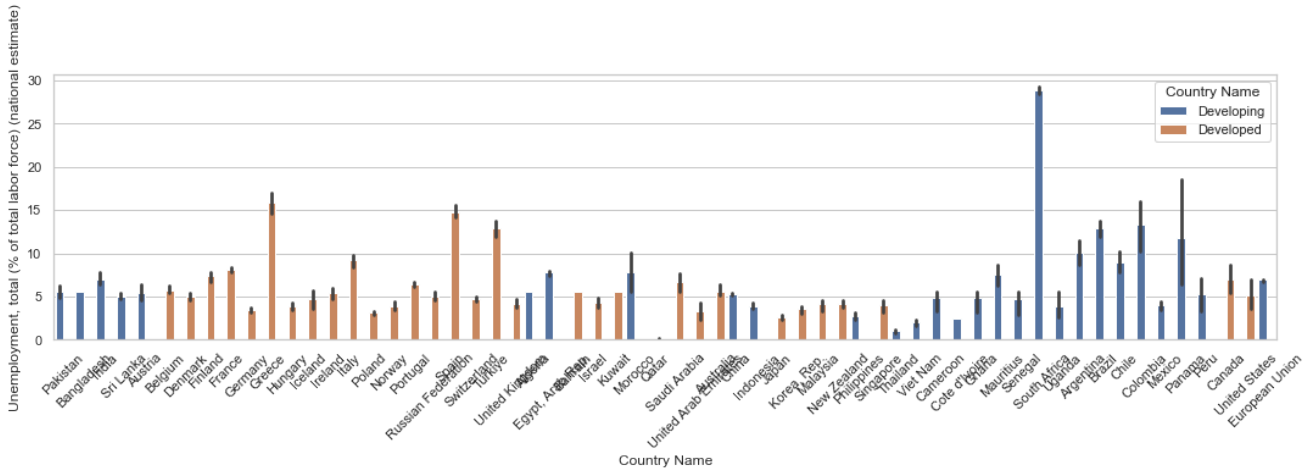

1e13
GNI (current US$)
Country Name
Developing
Developed
Country Name

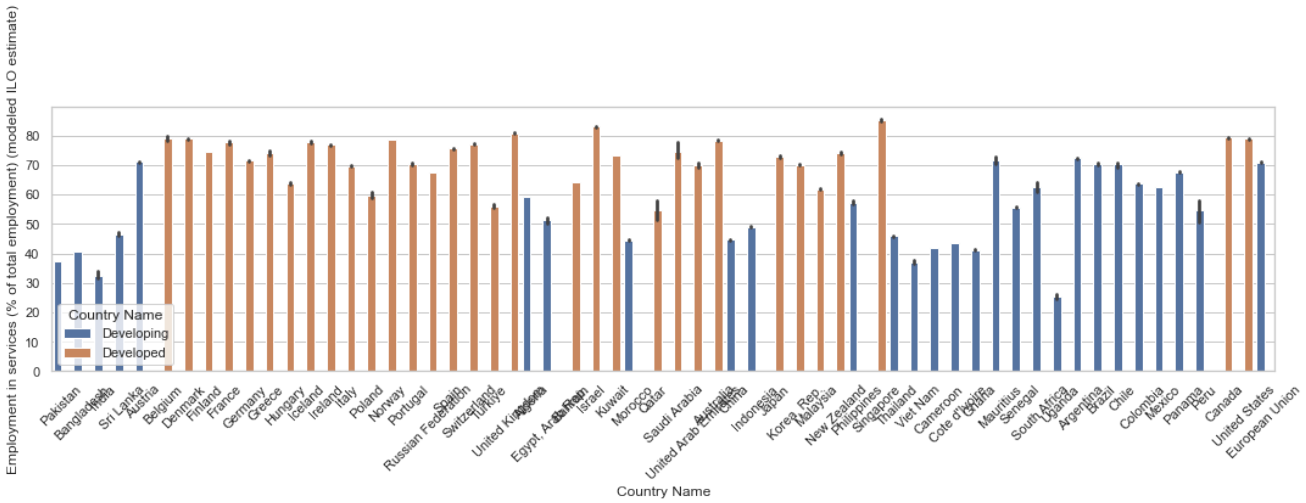

GNI growth (annual %)
Country Name
Developing
Developed
Country Name

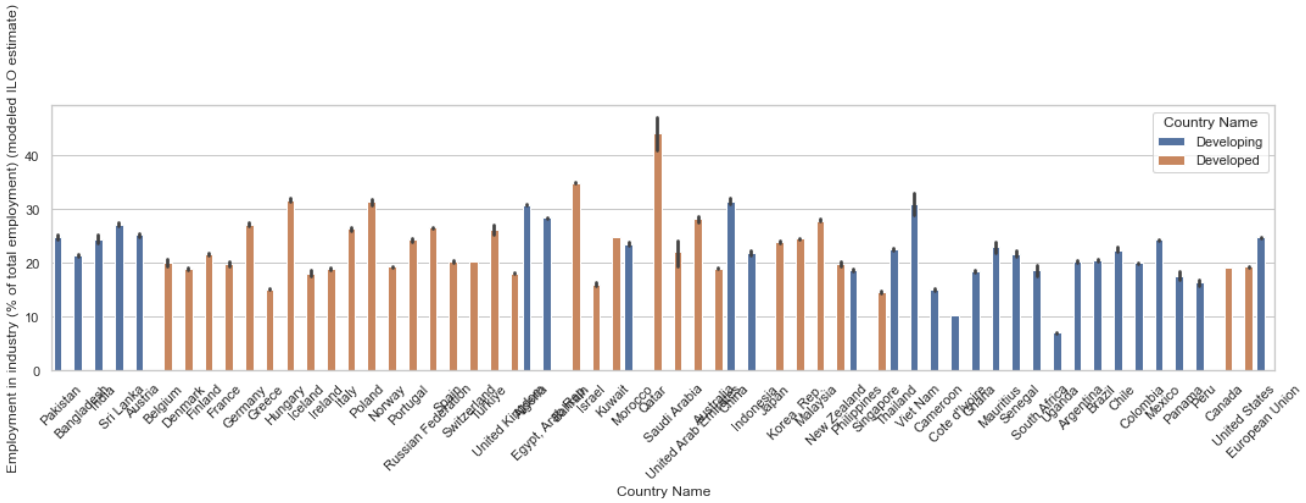

Employment to population ratio, 15+, total (%) (national estimate)
Country Name
Developing
Developed
Country Name

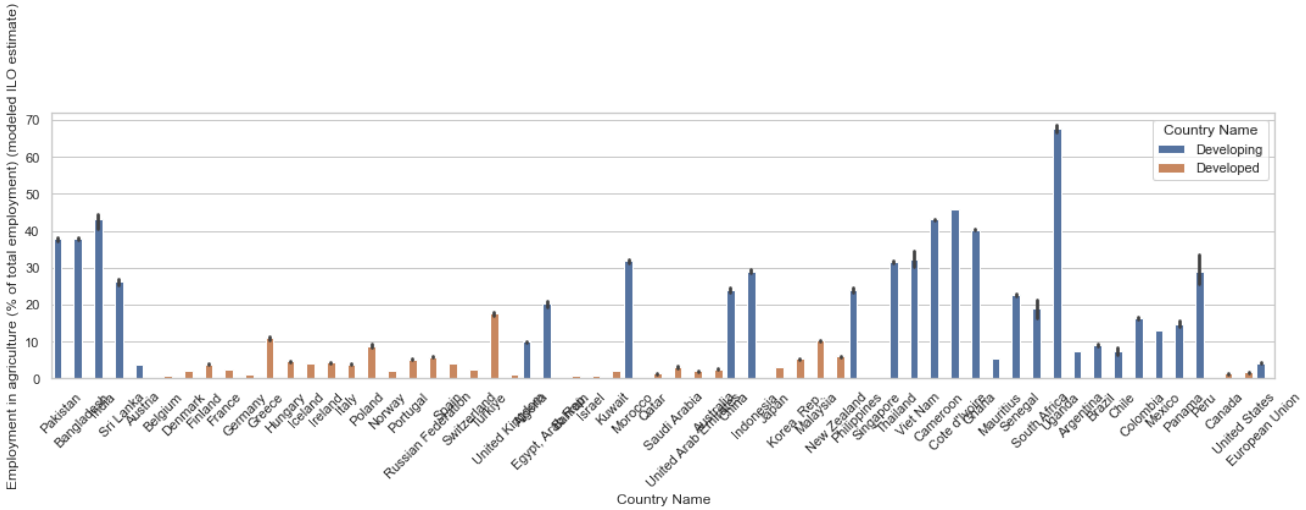

Country Name
Developing
Developed
Employment to population ratio, 15+, female (%) (national estimate)
Country Name


Country Name
Developing
Developed
Employment to population ratio, 15+, male (%) (national estimate)
Country Name


Country Name
Developing
Developed
Unemployment, total (% of total labor force) (national estimate)
Country Name

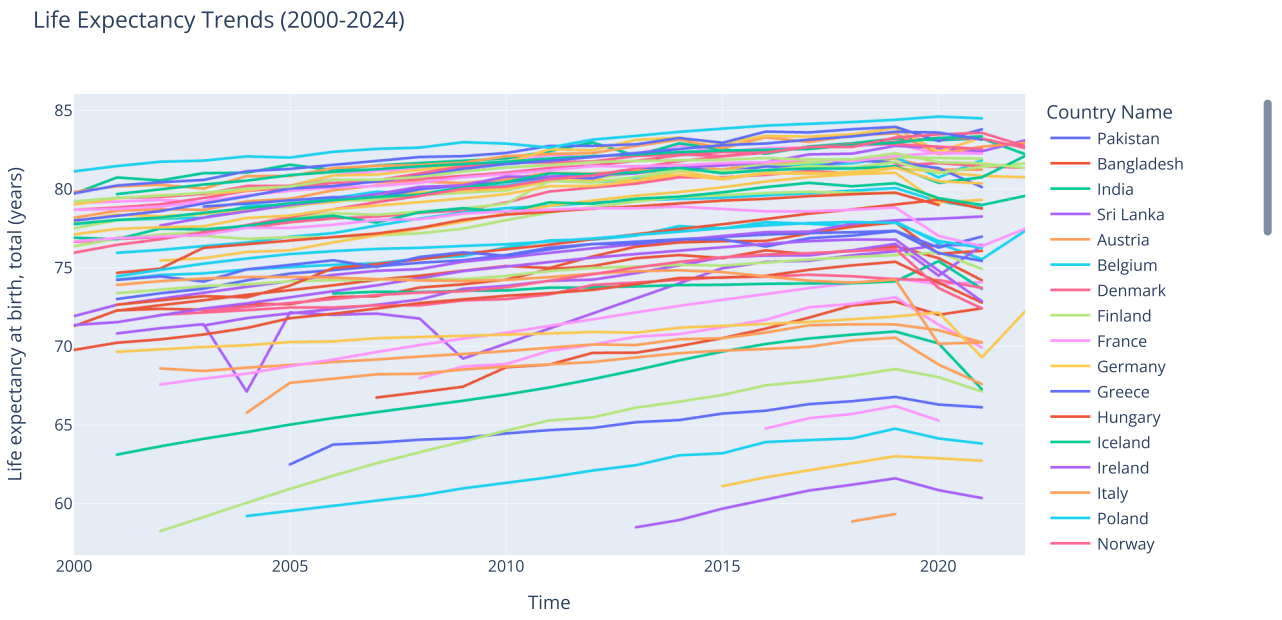




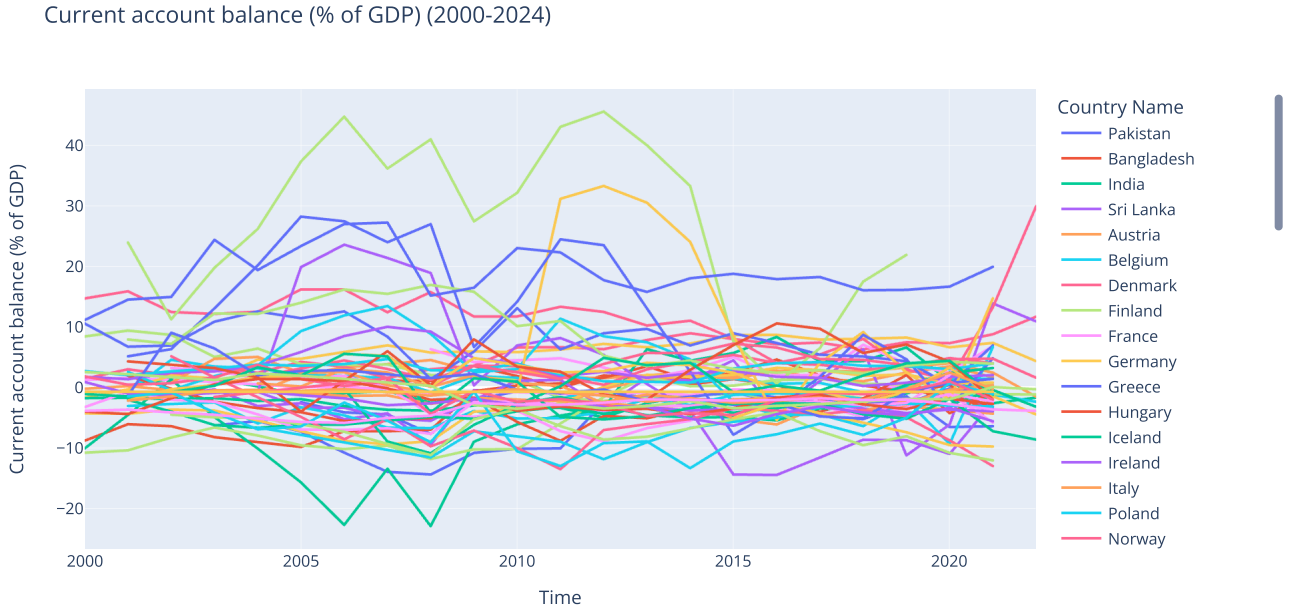


Bivariate Analysis

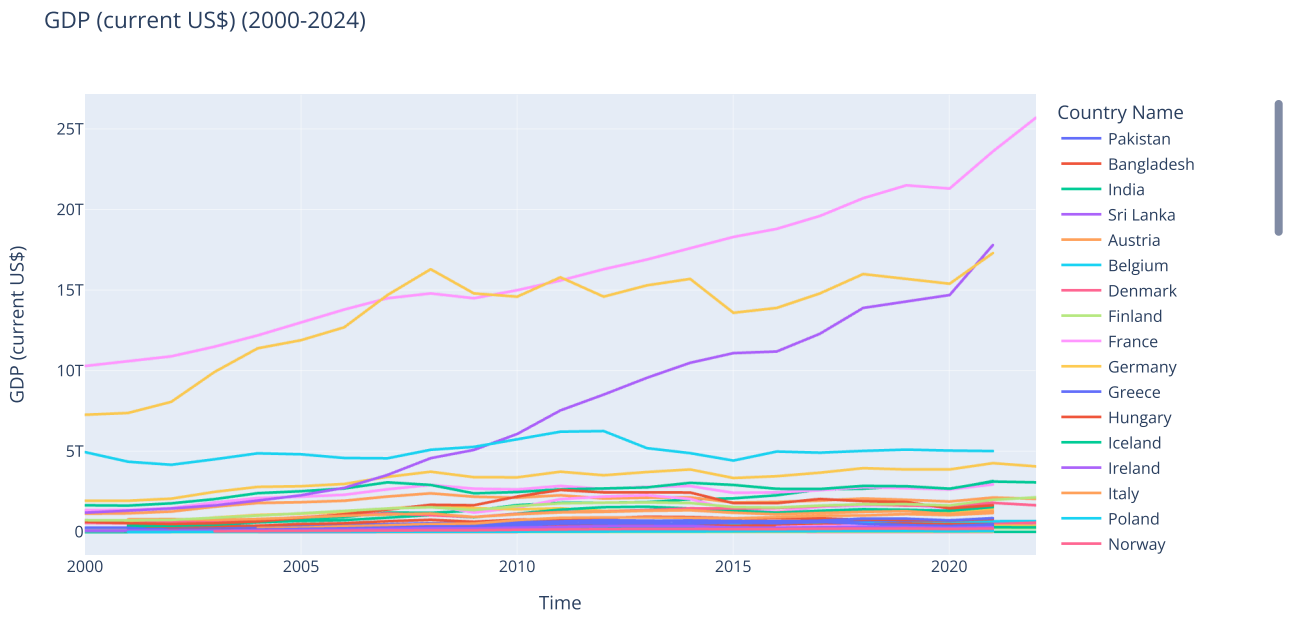
In [33]:
```
fig = px.line(df1, x='Time', y='Life expectancy at birth, total (years)', color='Country Name',
              title="Life Expectancy Trends (2000-2024)")
fig.show()
```



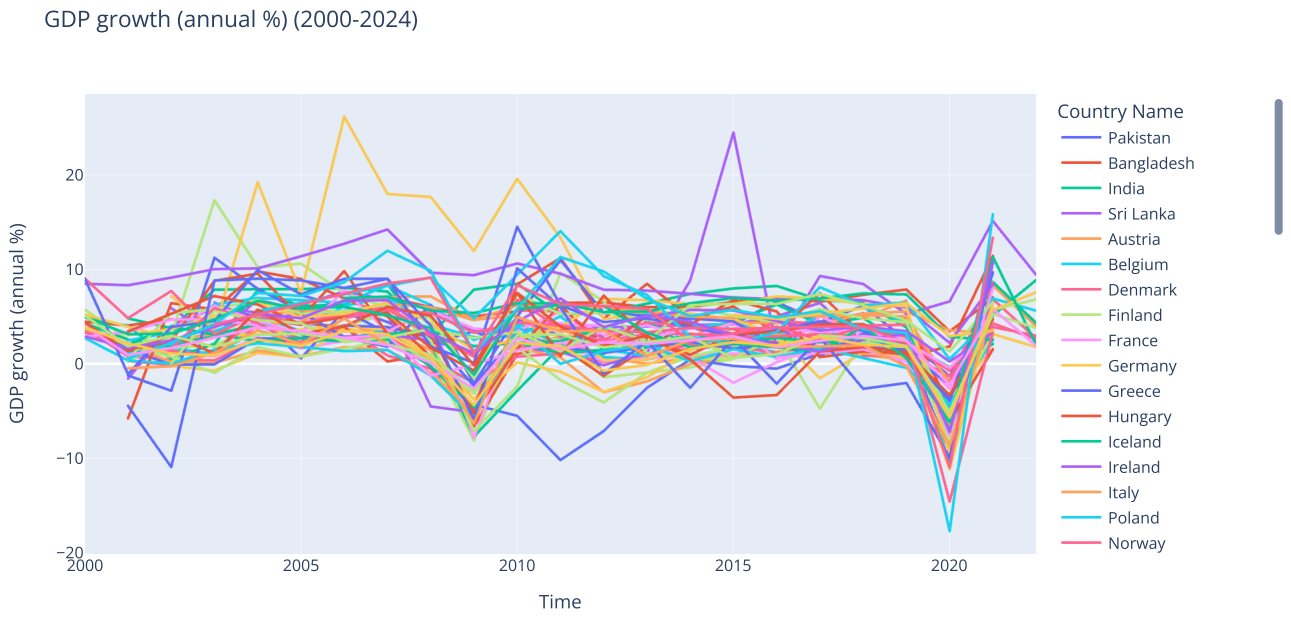
In [34]:
```
fig = px.line(df1, x='Time', y='Current account balance (% of GDP)', color='Country Name',
              title=f"Current account balance (% of GDP) (2000-2024)")
fig.show()
```



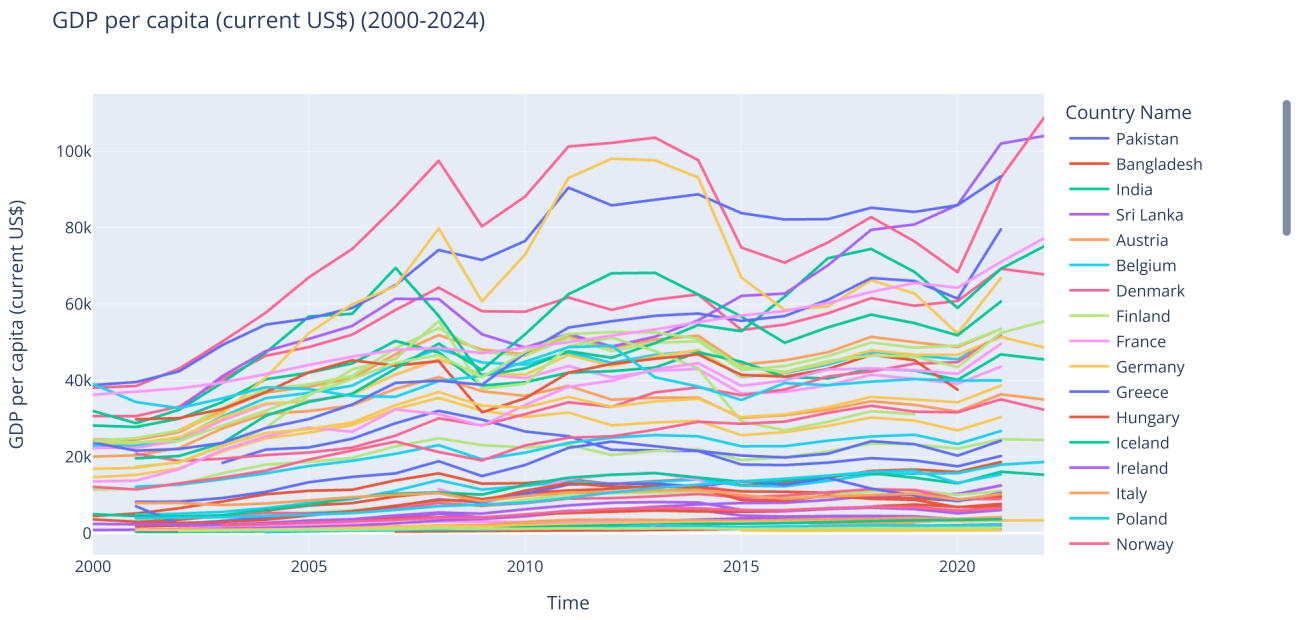
In [35]:
```
fig = px.line(df1, x='Time', y='GDP (current US$)', color='Country Name',
              title=f"GDP (current US$) (2000-2024)")
fig.show()
```



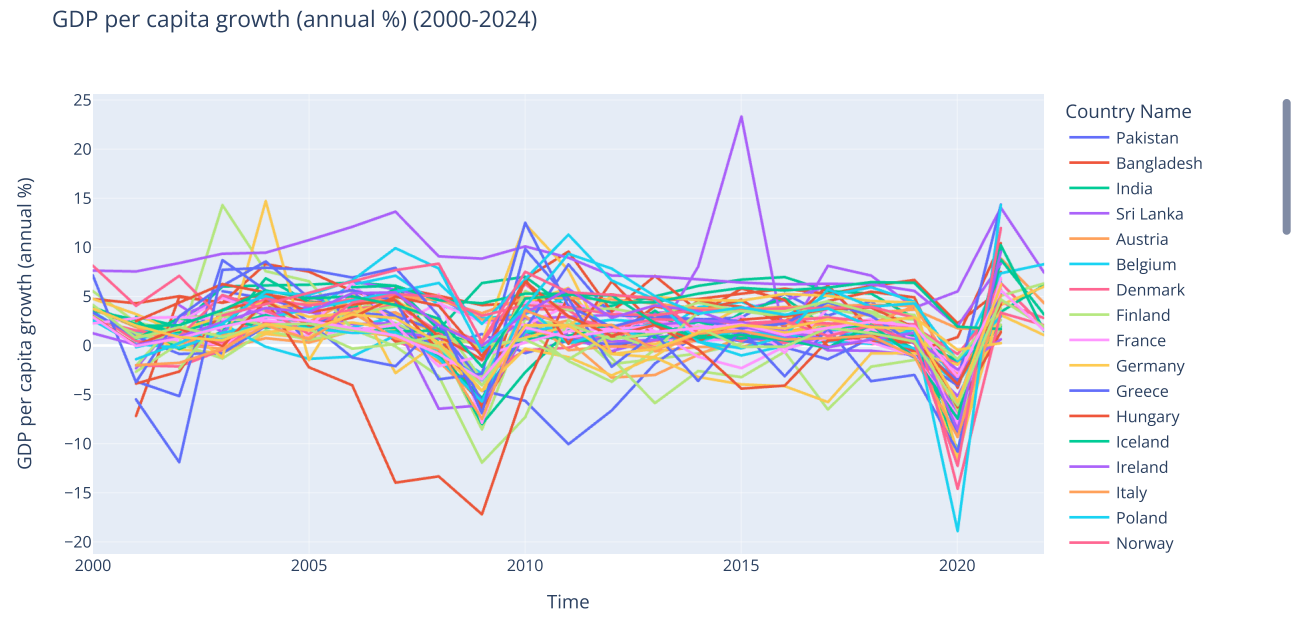
In [36]:
```
fig = px.line(df1, x='Time', y='GDP growth (annual %)', color='Country Name',
              title=f"GDP growth (annual %) (2000-2024)")
fig.show()
```



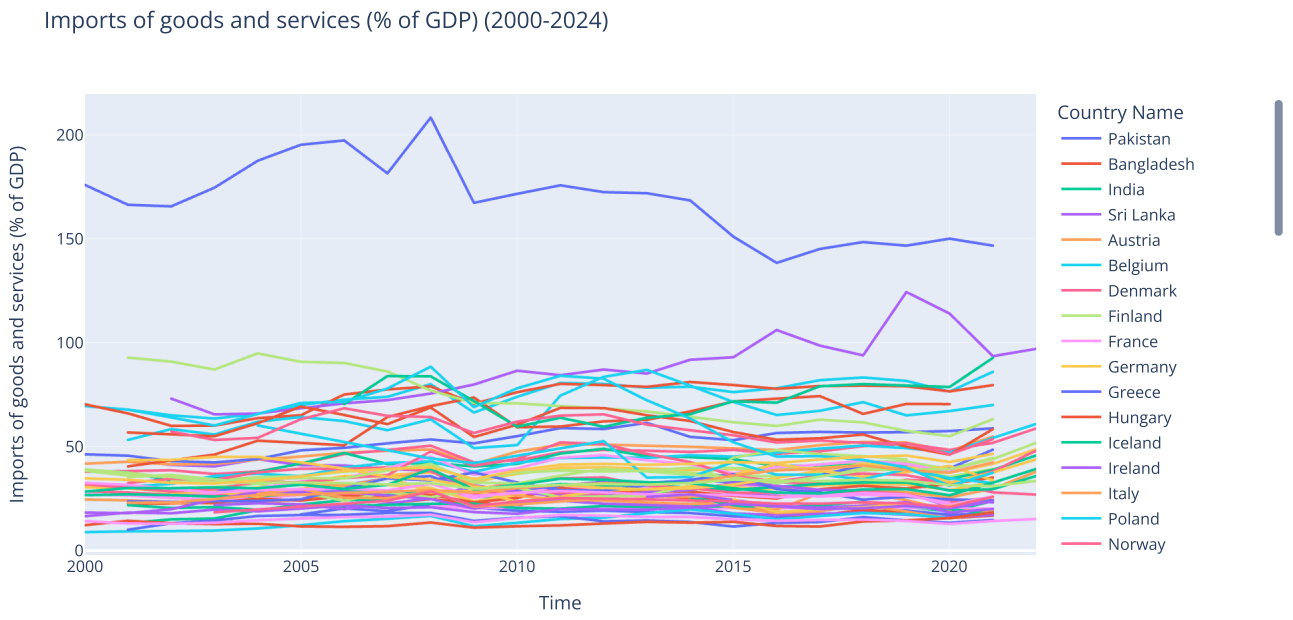
In [37]:
```
fig = px.line(df1, x='Time', y='GDP per capita (current US$)', color='Country Name',
              title=f"GDP per capita (current US$) (2000-2024)")
fig.show()
```



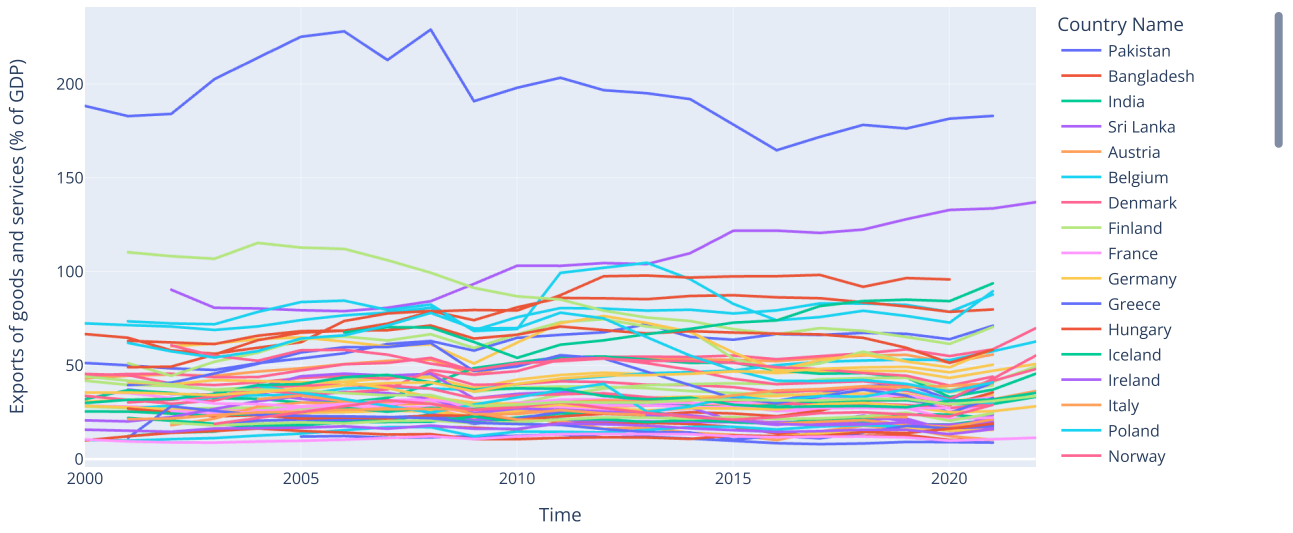
In [38]:
```
fig = px.line(df1, x='Time', y='GDP per capita growth (annual %)', color='Country Name',
              title=f"GDP per capita growth (annual %) (2000-2024)")
fig.show()
```



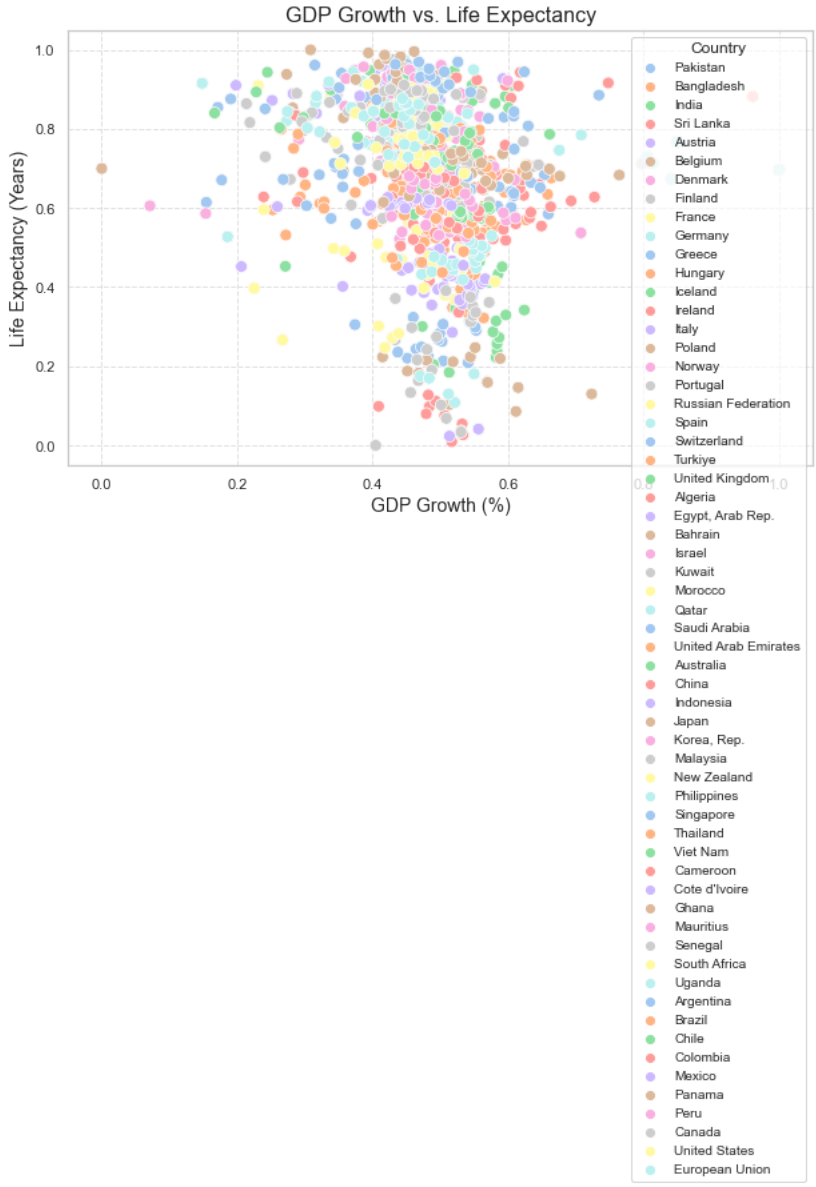
In [39]:
```
fig = px.line(df1, x='Time', y='Imports of goods and services (% of GDP)', color='Country Name',
              title=f"Imports of goods and services (% of GDP) (2000-2024)")
fig.show()
```



In [40]:
```
fig = px.line(df1, x='Time', y='Exports of goods and services (% of GDP)', color='Country Name',
              title=f"Exports of goods and services (% of GDP) (2000-2024)")
fig.show()
```

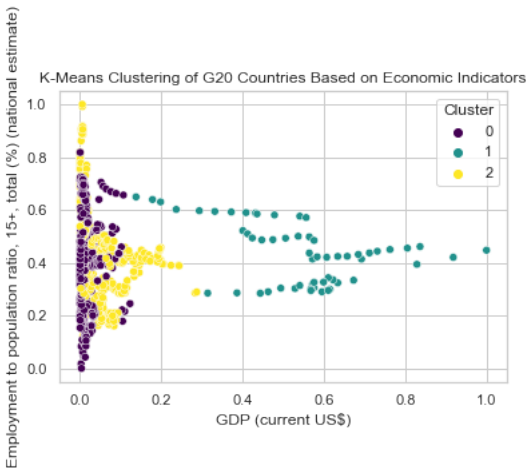


In [44]:
```
plt.figure(figsize=(10, 6))
sns.scatterplot(data=df, x='GDP growth (annual %)', y='Life expectancy at birth, total (years)', hue='Country Name', palette='pas
plt.title('GDP Growth vs. Life Expectancy', fontsize=16)
plt.xlabel('GDP Growth (%)', fontsize=14)
plt.ylabel('Life Expectancy (Years)', fontsize=14)
plt.legend(title='Country', fontsize=10)
plt.grid(axis='both', linestyle='--', alpha=0.5)
plt.show()
```

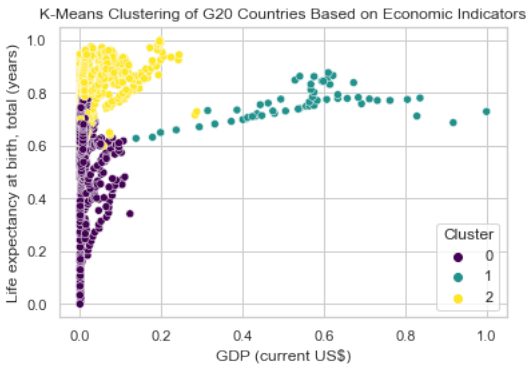


Mutlivariate Analysis

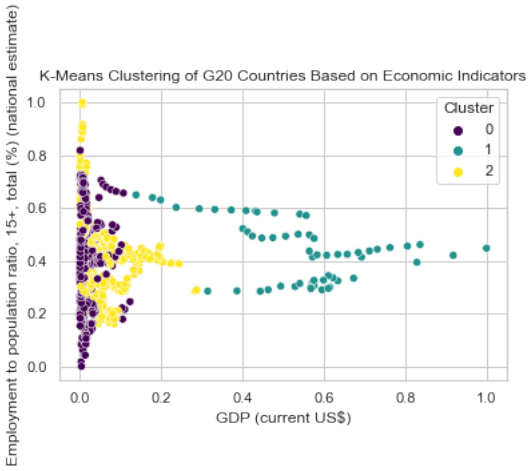
In [69]:
```
from sklearn.preprocessing import StandardScaler
from sklearn.cluster import KMeans

df_numeric = df.select_dtypes(["float64"])
kmeans = KMeans(n_clusters=3, random_state=42)
clusters = kmeans.fit_predict(df_numeric)

df['Cluster'] = clusters

sns.scatterplot(x=df['GDP (current US$)'], y=df['Employment to population ratio, 15+, total (%) (national estimate)'], hue=df['Cl
plt.title('K-Means Clustering of G20 Countries Based on Economic Indicators')
plt.show()
```



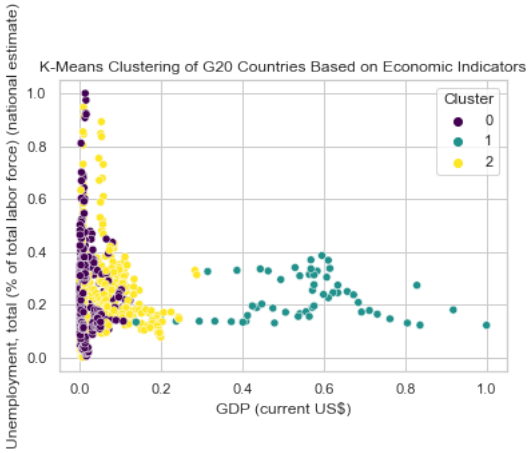
In [70]:
```
sns.scatterplot(x=df['GDP (current US$)'], y=df['Life expectancy at birth, total (years)'], hue=df['Cluster'], palette="viridis")
plt.title('K-Means Clustering of G20 Countries Based on Economic Indicators')
plt.show()
```



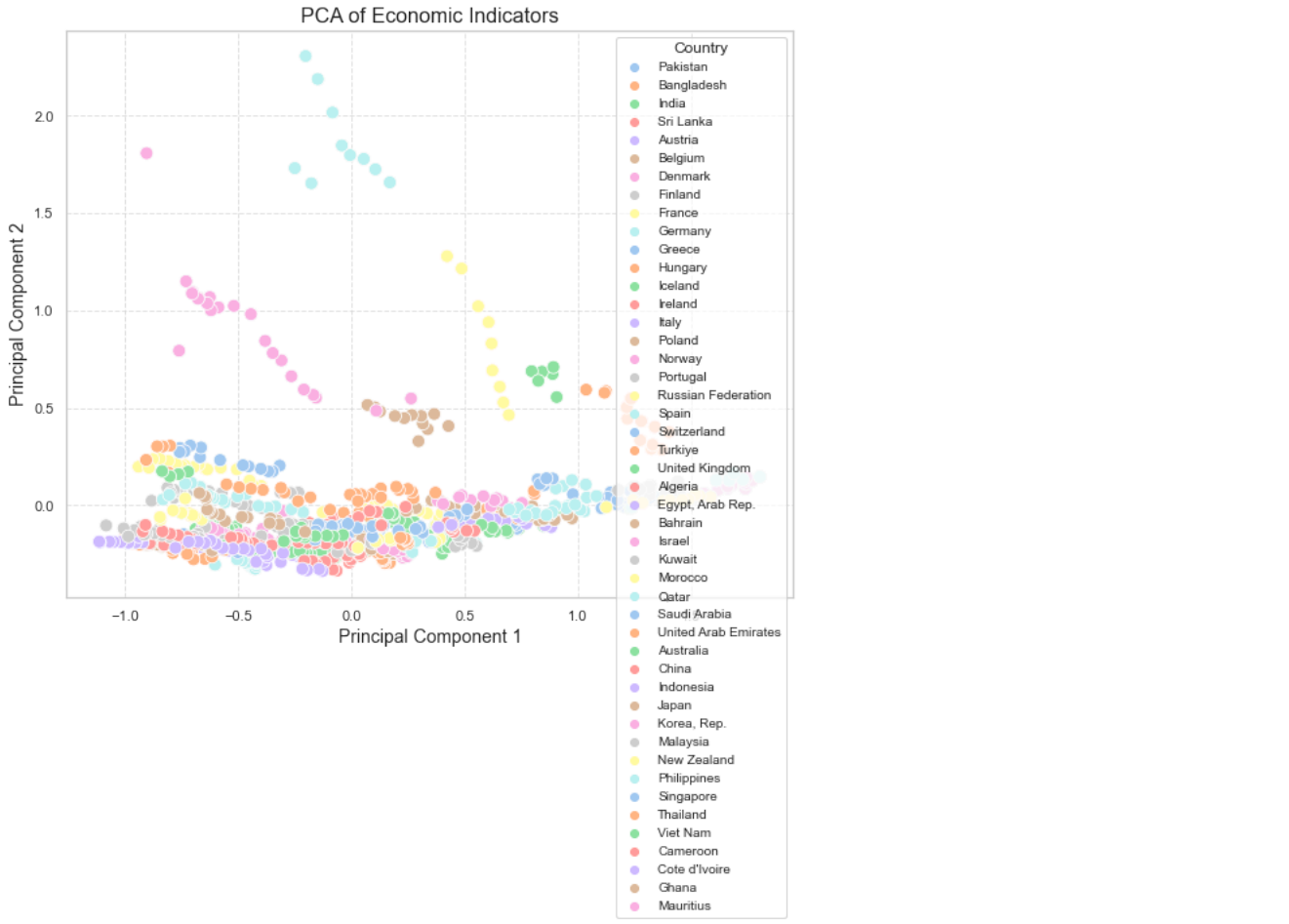
In [71]:
```
sns.scatterplot(x=df['GDP (current US$)'], y=df['Employment to population ratio, 15+, total (%) (national estimate)'], hue=df['Cl
plt.title('K-Means Clustering of G20 Countries Based on Economic Indicators')
plt.show()
```



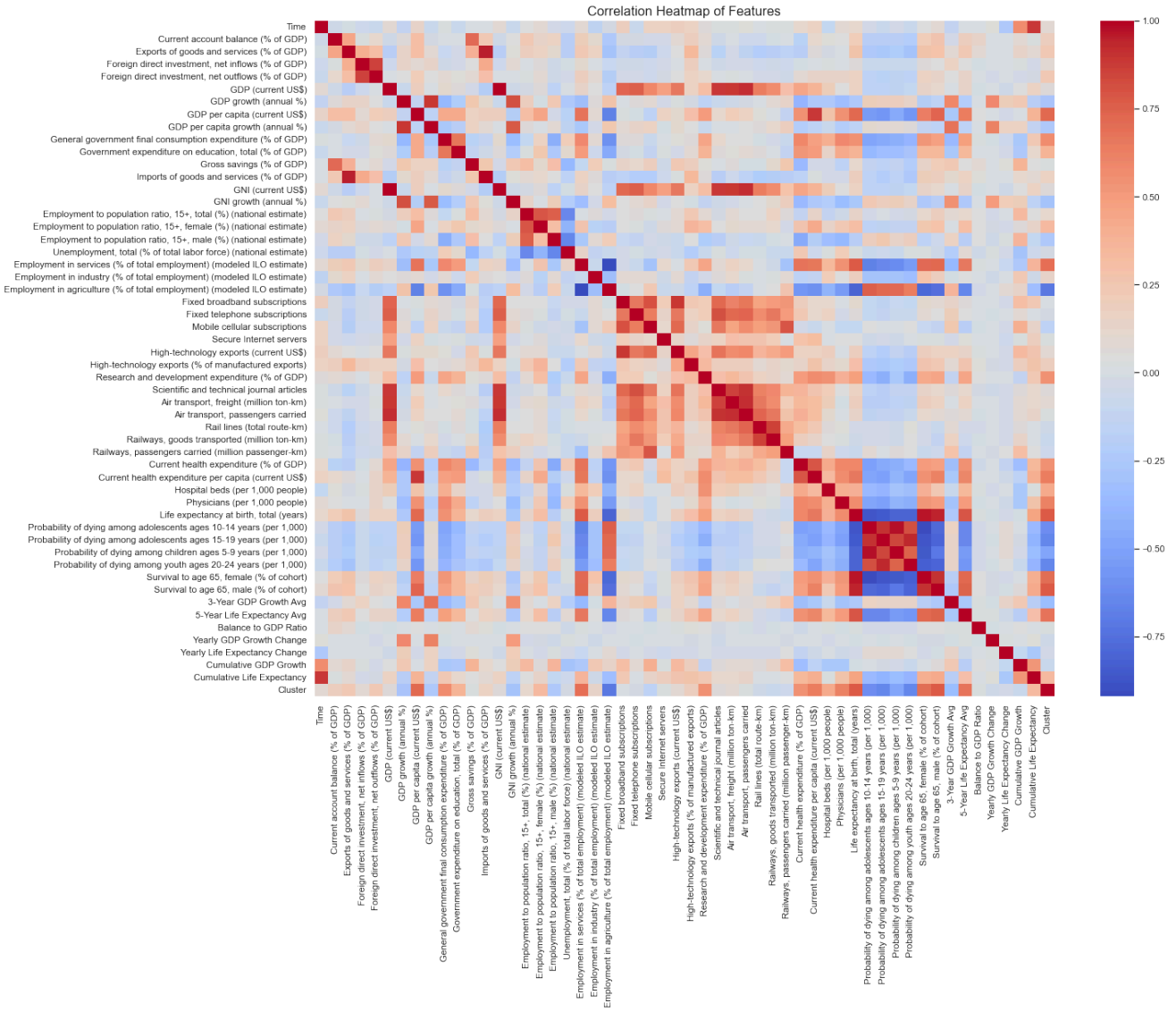
In [72]:
```
sns.scatterplot(x=df['GDP (current US$)'], y=df['Unemployment, total (% of total labor force) (national estimate)'], hue=df['Clus
plt.title('K-Means Clustering of G20 Countries Based on Economic Indicators')
plt.show()
```



In [59]:
```
from sklearn.decomposition import PCA

pca = PCA(n_components=2)
pca_result = pca.fit_transform(df[scaled_columns])
pca_df = pd.DataFrame(pca_result, columns=['PC1', 'PC2'])
pca_df['Country'] = df['Country Name']

plt.figure(figsize=(10, 8))
sns.scatterplot(data=pca_df, x='PC1', y='PC2', hue='Country', palette='pastel', s=100)
plt.title('PCA of Economic Indicators', fontsize=16)
plt.xlabel('Principal Component 1', fontsize=14)
plt.ylabel('Principal Component 2', fontsize=14)
plt.legend(title='Country', fontsize=10)
plt.grid(axis='both', linestyle='--', alpha=0.5)
plt.show()
```



In [65]:
```
plt.figure(figsize=(20, 15))
correlation_matrix = df.corr()
sns.heatmap(correlation_matrix, cmap='coolwarm')
plt.title('Correlation Heatmap of Features', fontsize=16)
plt.show()
```



Case Study: 2008 Global Economic Crisis

The 2007–2008 financial crisis, or the global financial crisis (GFC), was the most severe worldwide economic crisis since the 1929 Wall Street crash that began the Great Depression. Causes of the crisis included predatory lending in the form of subprime mortgages to low-income homebuyers and a resulting housing bubble, excessive risk-taking by global financial institutions, and lack of regulatory oversight, which culminated in a "perfect storm" that triggered the Great Recession, which lasted from late 2007 to mid-2009. The financial crisis began in early 2007, as mortgage-backed securities (MBS) tied to U.S. real estate, as well as a vast web of derivatives linked to those MBS, collapsed in value. Financial institutions worldwide suffered severe damage, reaching a climax with the bankruptcy of Lehman Brothers on September 15, 2008, and a subsequent international banking crisis.

The prerequisites for the crisis were complex. During the 1990s, the U.S. Congress had passed legislation intended to expand affordable housing through looser financing. In 1999, parts of the Glass–Steagall legislation (passed in 1933) were repealed, permitting institutions to mix low-risk operations, such as commercial banking and insurance, with higher-risk operations such as investment banking and proprietary trading. As the Federal Reserve ("Fed") lowered the federal funds rate from 2000 to 2003, institutions increasingly targeted low-income homebuyers, largely belonging to racial minorities, with high-risk loans; this development went unattended by regulators. As interest rates rose from 2004 to 2006, the cost of mortgages rose and the demand for housing fell, causing property values to decline. In early 2007, as more U.S. mortgage holders began defaulting on their repayments, subprime lenders went bankrupt, culminating in April with the bankruptcy of New Century Financial. As demand and prices continued to fall, the contagion spread to worldwide credit markets by August, and central banks began injecting liquidity. By July 2008, Fannie Mae and Freddie Mac, companies which together owned or guaranteed half of the U.S. housing market, were on the verge of collapse; the Housing and Economic Recovery Act enabled the government to take over and cover their combined $1.6 trillion debt on September 7.

For this analysis we will focus only on the G20 nations

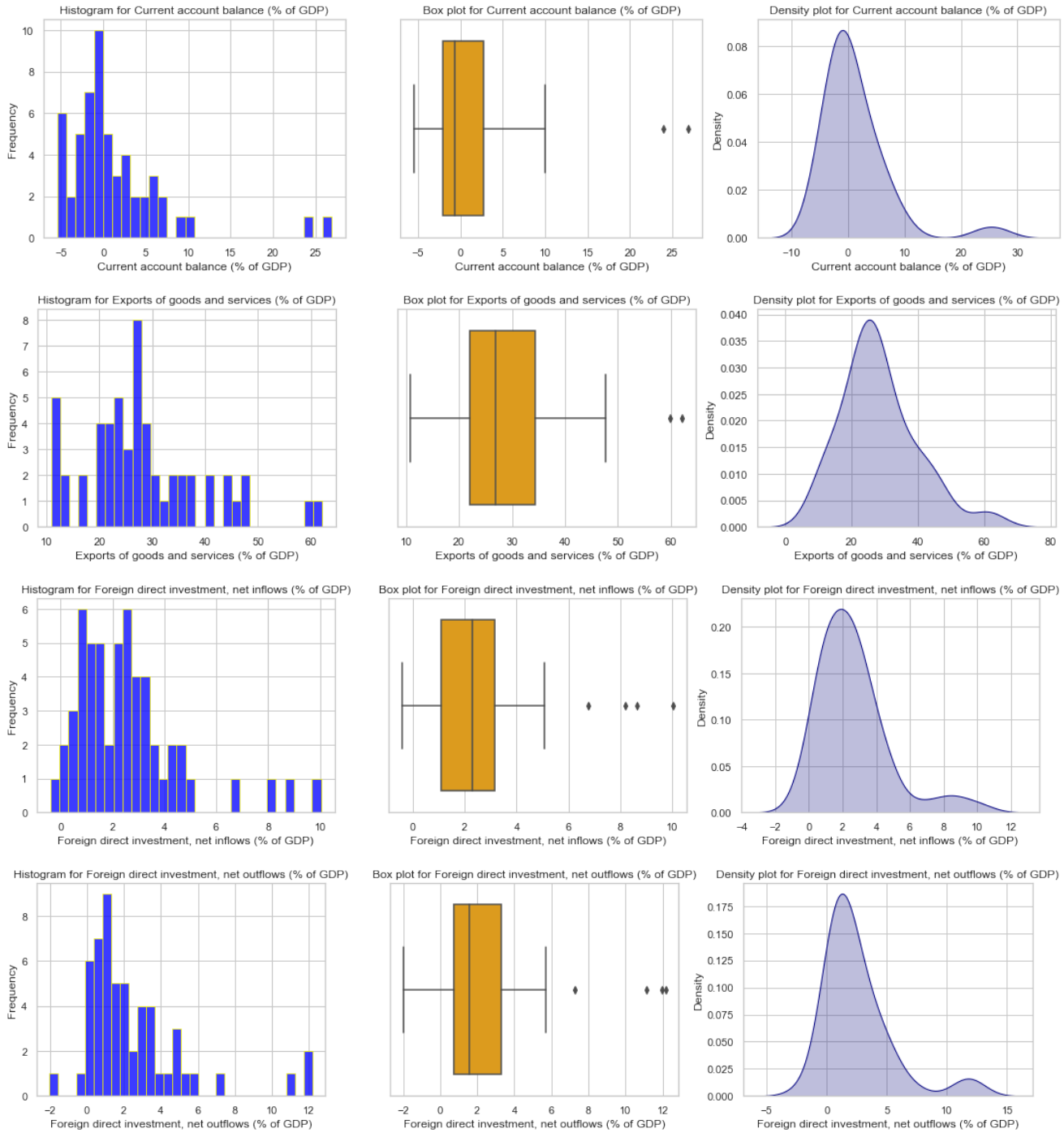
In [87]:
```
g20 = ['Argentina', 'Australia', 'Brazil', 'Canada', 'China', 'France', 'Germany', 'India', 'Indonesia', 'Italy', 'Japan', 'Mexic
for i in g20:
    if i not in df["Country Name"].unique():
        print(i)

years = [2007, 2008, 2009]

df_economic_crisis = df1[(df1["Country Name"].isin(g20)) & (df1["Time"].isin(years))]
df_economic_crisis.head()
```

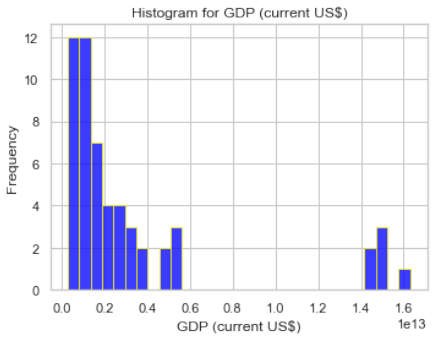
Out[87]:

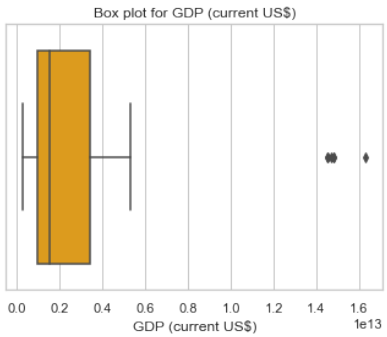
	Country Name	Country Code	Time	Current account balance (% of GDP)	Exports of goods and services (% of GDP)	Foreign direct investment, net inflows (% of GDP)	Foreign direct investment, net outflows (% of GDP)	GDP (current US$)	GDP growth (annual %)	GDP per capita (current US$)	...	Current health expenditure per capita (current US$)	Hospital beds (per 1,000 people)	Physicians (per 1,000 people)	expe a
53	India	IND	2007	-0.663718	20.799700	2.073394	1.399326	1.220000e+12	7.660815	1022.732467	...	36.000000	1.90	2.289	65.
54	India	IND	2008	-2.583377	24.097357	3.620523	1.606189	1.200000e+12	3.086698	993.503405	...	38.000000	1.87	2.289	66.
55	India	IND	2009	-1.951462	20.400519	2.651590	1.199473	1.340000e+12	7.861889	1096.636136	...	38.000000	1.84	2.289	66.
191	France	FRA	2007	-0.327691	27.853258	3.148834	4.969835	2.660000e+12	2.424736	41557.623560	...	4312.310547	7.35	3.449	81.
192	France	FRA	2008	-0.961095	28.119385	2.320491	4.641131	2.930000e+12	0.254946	45515.961750	...	4807.374512	7.18	3.443	81.

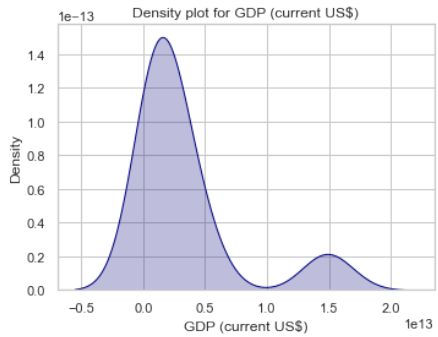
5 rows × 48 columns

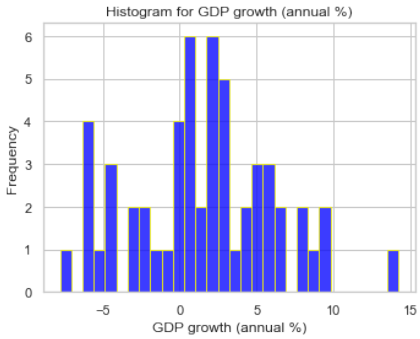
In [88]:
```
for i in economic_metrics:
    plt.figure(figsize=(15,4))

    plt.subplot(1,3,1)
    sns.histplot(data=df_economic_crisis[i], color="blue", edgecolor="yellow", bins=30)
    plt.xlabel(i)
    plt.ylabel("Frequency")
    plt.title(f"Histogram for {i}")

    plt.subplot(1,3,2)
    sns.boxplot(x=df_economic_crisis[i], color="orange")
    plt.xlabel(i)
    plt.title(f"Box plot for {i}")

    plt.subplot(1,3,3)
    sns.kdeplot(data=df_economic_crisis[i], color="navy", fill=True)
    plt.xlabel(i)
    plt.ylabel("Density")
    plt.title(f"Density plot for {i}")

    plt.tight_layout()
    plt.show()
```

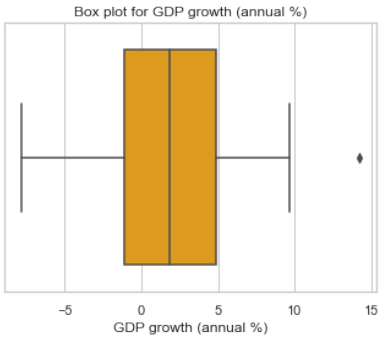


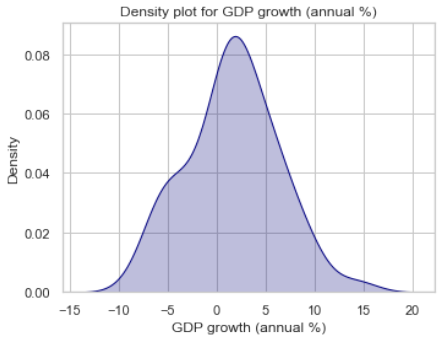

Histogram for GDP (current US$)
Frequency
GDP (current US$)
1e13

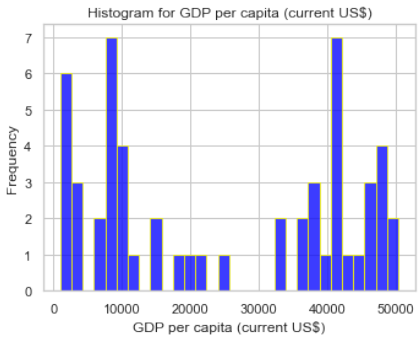

Box plot for GDP (current US$)
GDP (current US$)
1e13

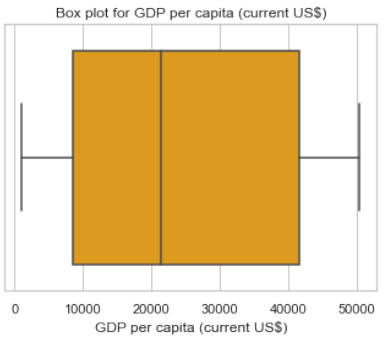

Density plot for GDP (current US$)
1e−13
Density
GDP (current US$)
1e13

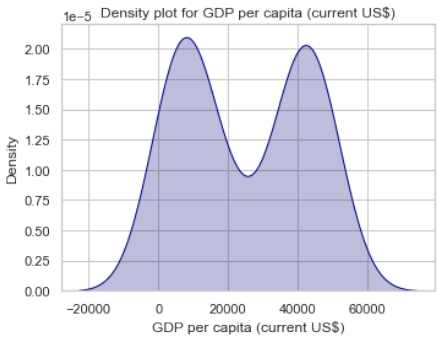

Histogram for GDP growth (annual %)
Frequency
GDP growth (annual %)

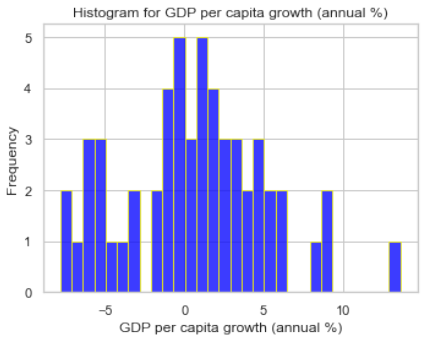

Box plot for GDP growth (annual %)
GDP growth (annual %)

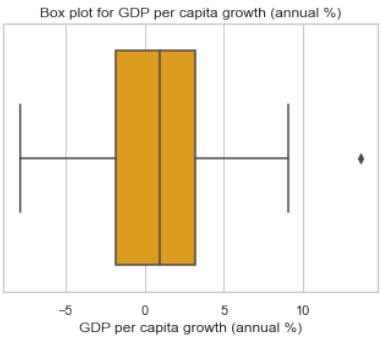

Density plot for GDP growth (annual %)
Density
GDP growth (annual %)

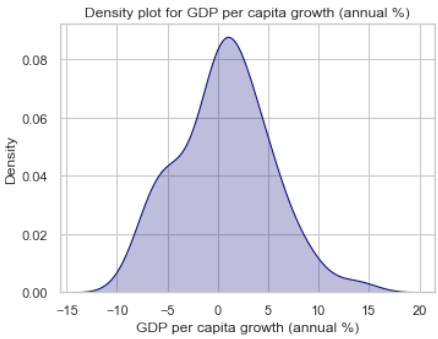

Histogram for GDP per capita (current US$)
Frequency
GDP per capita (current US$)

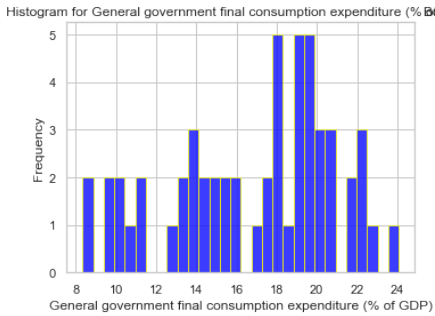

Box plot for GDP per capita (current US$)
GDP per capita (current US$)

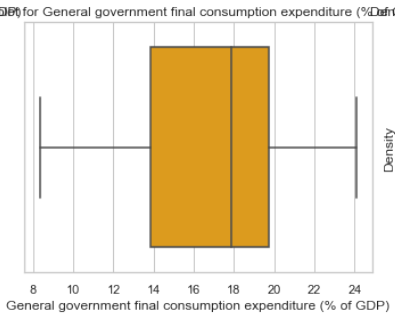

1e−5
Density plot for GDP per capita (current US$)
Density
GDP per capita (current US$)

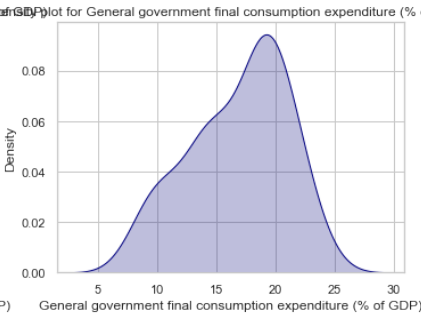

Histogram for GDP per capita growth (annual %)
Frequency
GDP per capita growth (annual %)

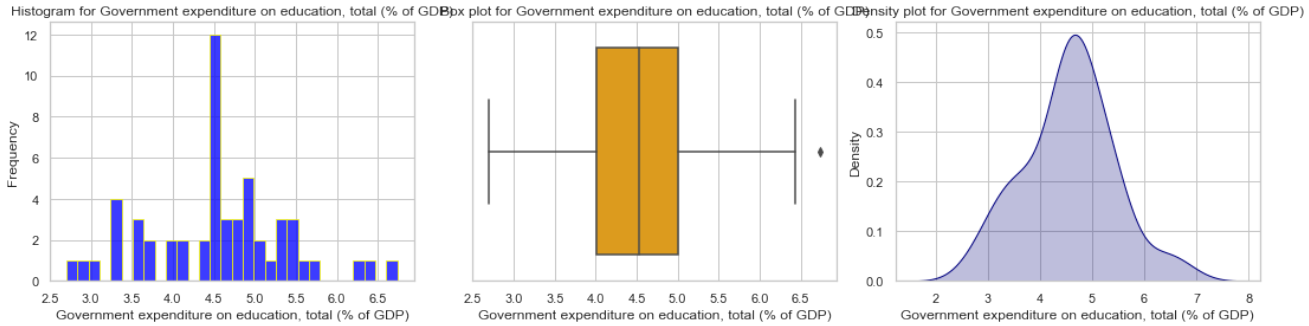

Box plot for GDP per capita growth (annual %)
GDP per capita growth (annual %)

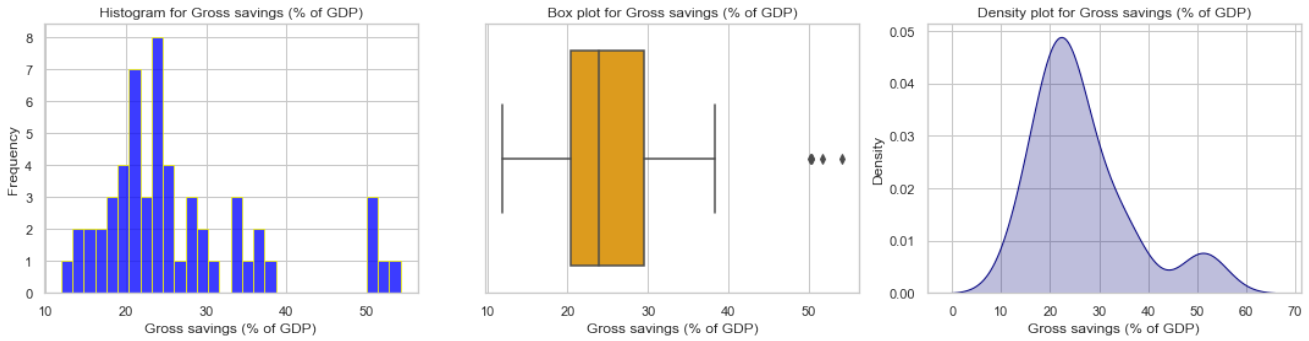

Density plot for GDP per capita growth (annual %)
Density
GDP per capita growth (annual %)

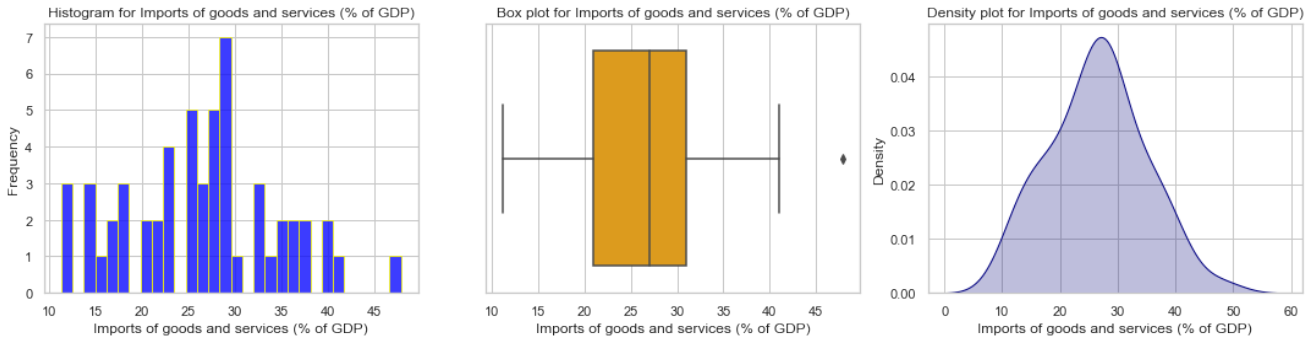

Histogram for General government final consumption expenditure (% of GDP)
Frequency
General government final consumption expenditure (% of GDP)

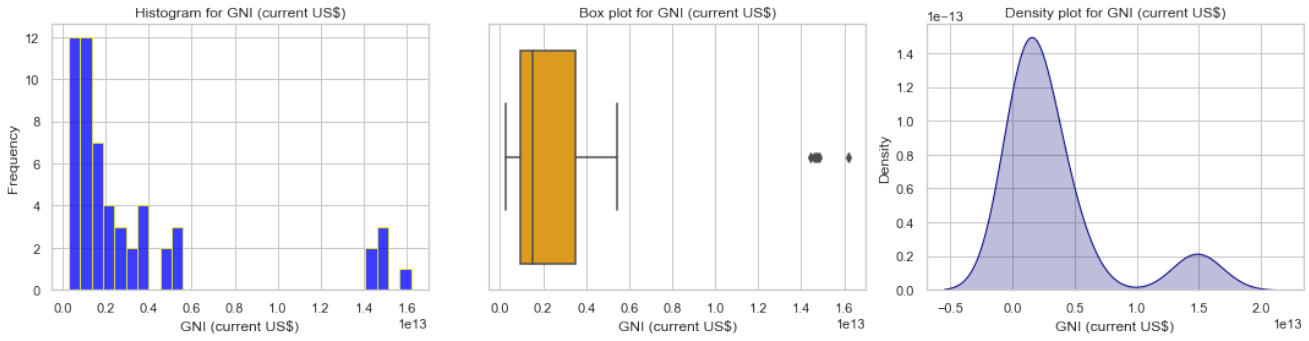

Box plot for General government final consumption expenditure (% of GDP)
General government final consumption expenditure (% of GDP)

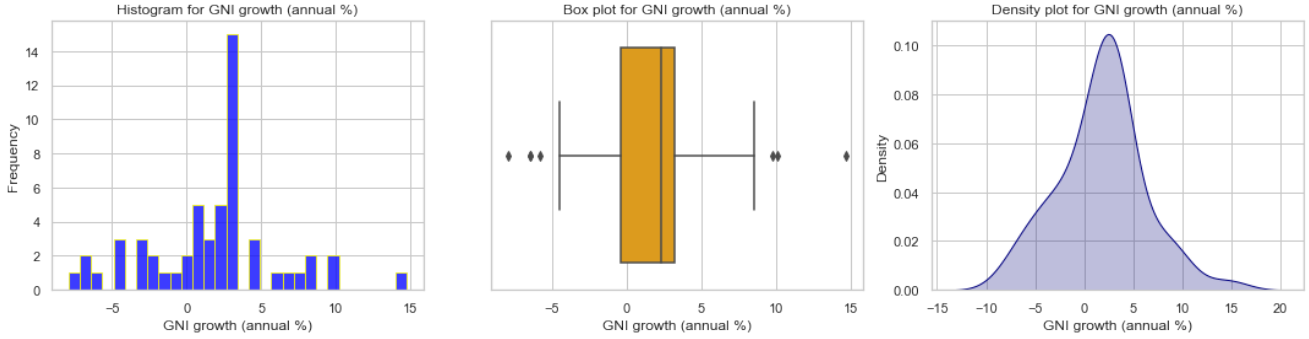

Density plot for General government final consumption expenditure (% of GDP)
Density
General government final consumption expenditure (% of GDP)

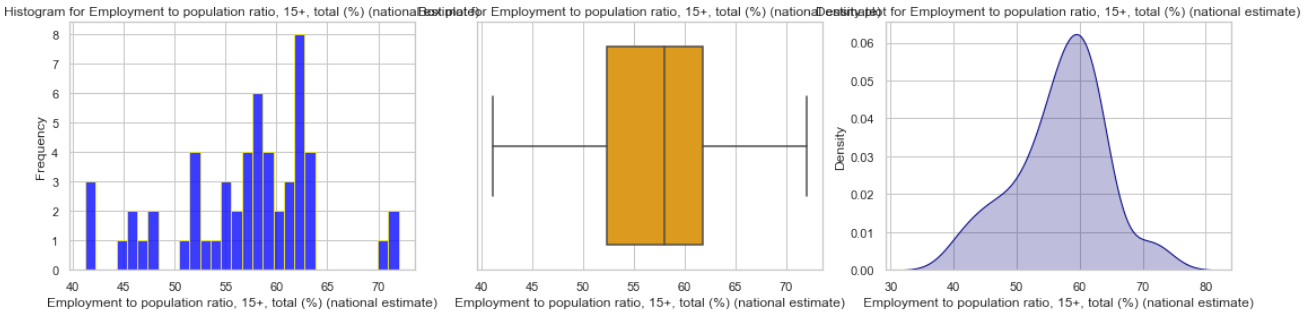

Histogram for Government expenditure on education, total (% of GDP)
Box plot for Government expenditure on education, total (% of GDP)
Density plot for Government expenditure on education, total (% of GDP)
Frequency
Density
Government expenditure on education, total (% of GDP)

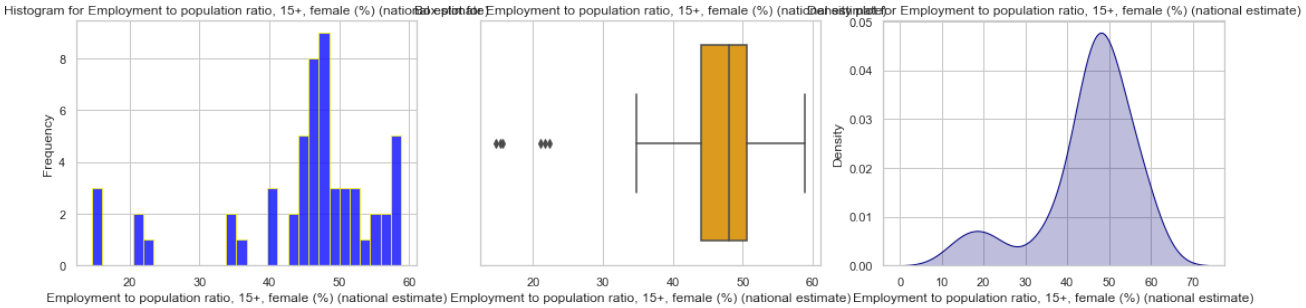

Histogram for Gross savings (% of GDP)
Box plot for Gross savings (% of GDP)
Density plot for Gross savings (% of GDP)
Frequency
Density
Gross savings (% of GDP)

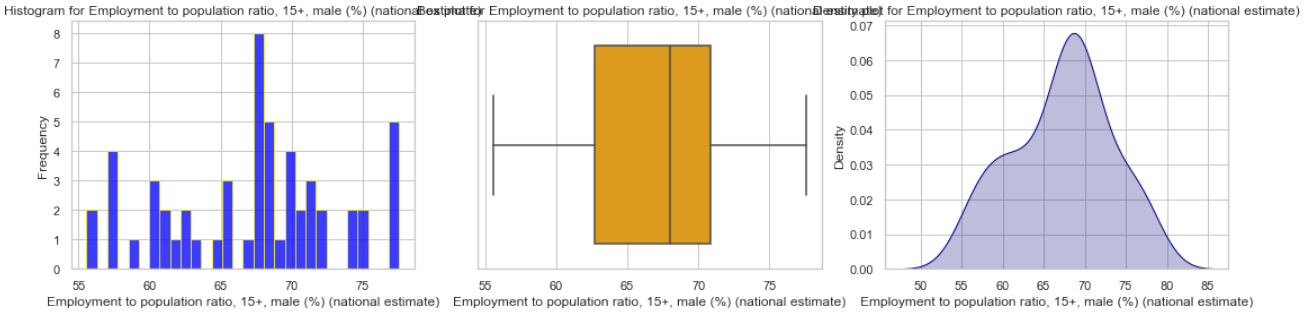

Histogram for Imports of goods and services (% of GDP)
Box plot for Imports of goods and services (% of GDP)
Density plot for Imports of goods and services (% of GDP)
Frequency
Density
Imports of goods and services (% of GDP)

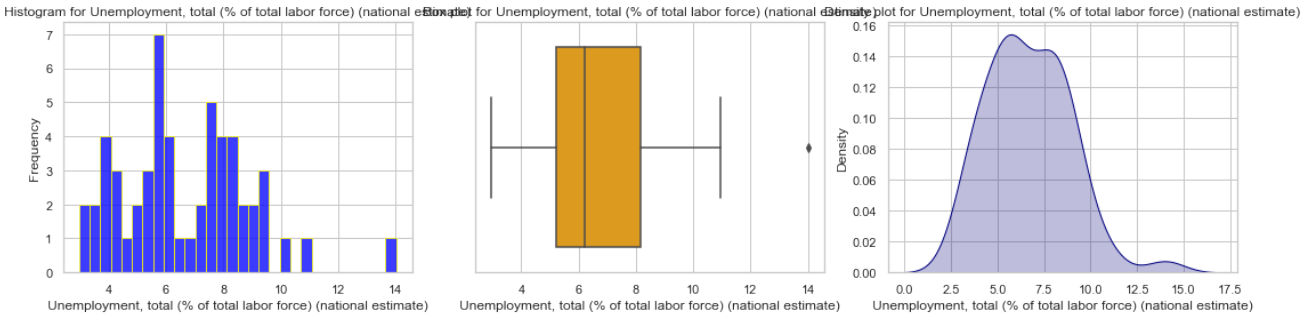

Histogram for GNI (current US$)
Box plot for GNI (current US$)
Density plot for GNI (current US$)
Frequency
Density
GNI (current US$)
1e13

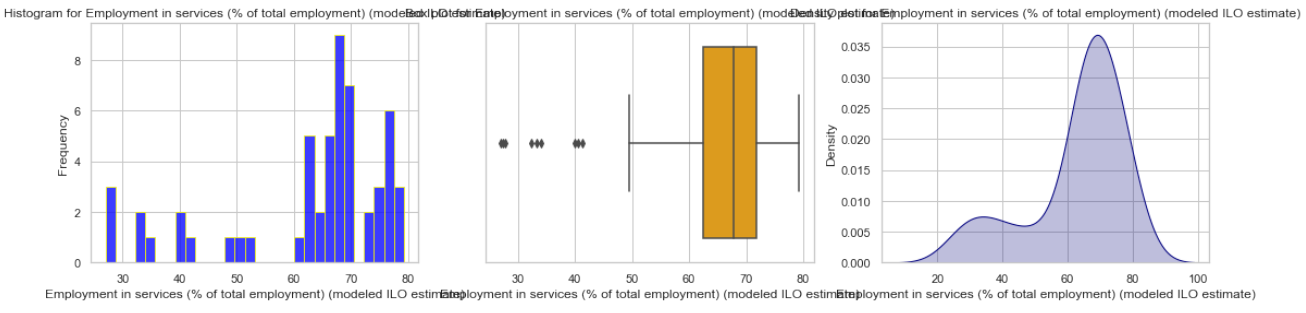

Histogram for GNI growth (annual %)
Box plot for GNI growth (annual %)
Density plot for GNI growth (annual %)
Frequency
Density
GNI growth (annual %)

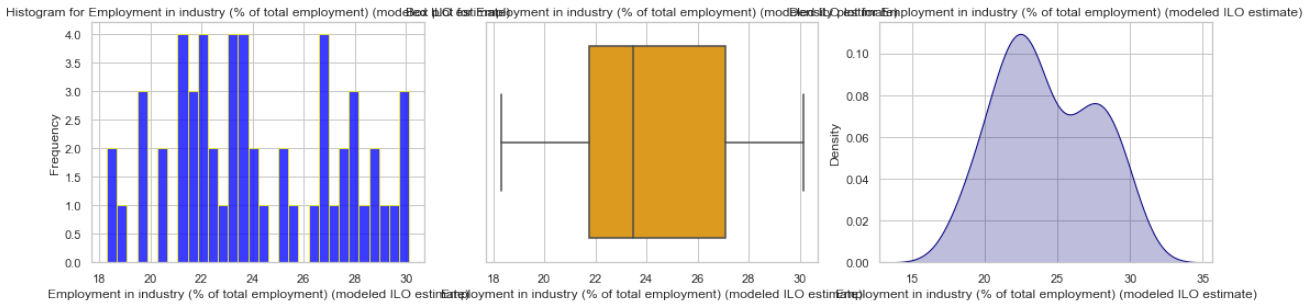

Histogram for Employment to population ratio, 15+, total (%) (national estimate)
Boxplot for Employment to population ratio, 15+, total (%) (national estimate)
Density plot for Employment to population ratio, 15+, total (%) (national estimate)
Frequency
Density
Employment to population ratio, 15+, total (%) (national estimate)

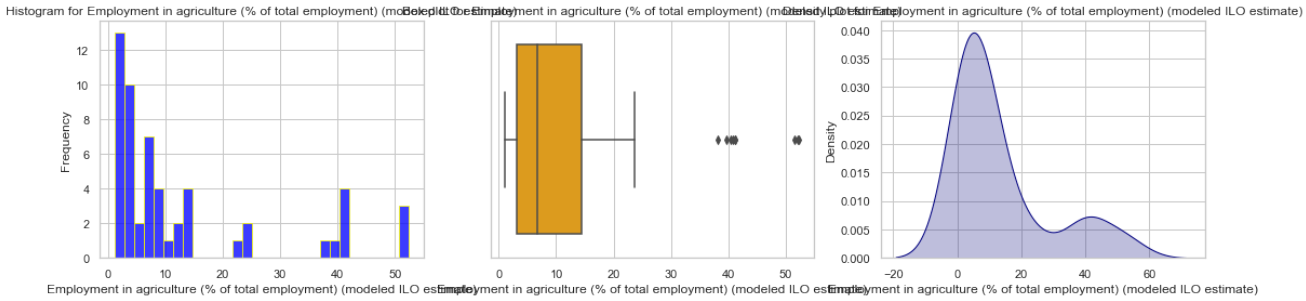

Histogram for Employment to population ratio, 15+, female (%) (national estimate)
Boxplot for Employment to population ratio, 15+, female (%) (national estimate)
Density plot for Employment to population ratio, 15+, female (%) (national estimate)
Frequency
Density
Employment to population ratio, 15+, female (%) (national estimate)


Histogram for Employment to population ratio, 15+, male (%) (national estimate)
Boxplot for Employment to population ratio, 15+, male (%) (national estimate)
Density plot for Employment to population ratio, 15+, male (%) (national estimate)
Frequency
Density
Employment to population ratio, 15+, male (%) (national estimate)

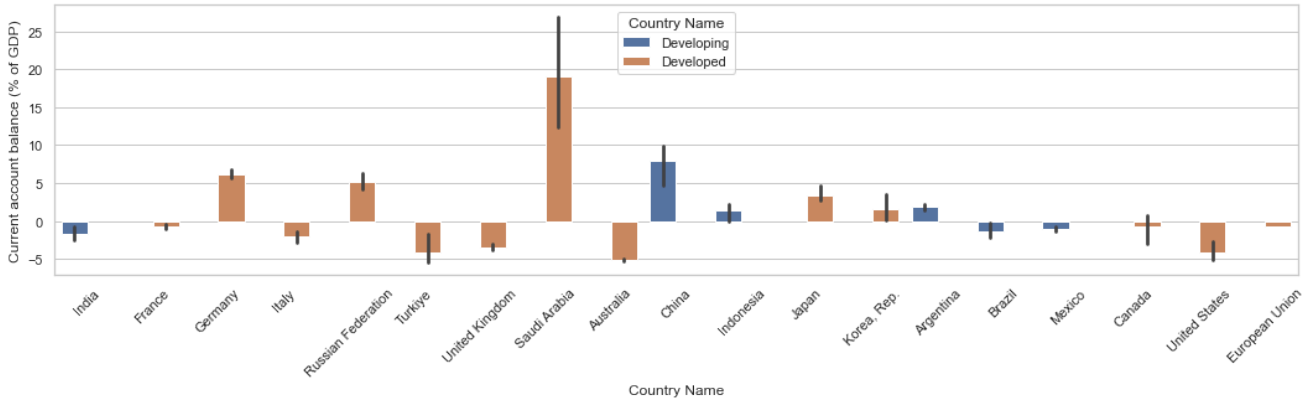

Histogram for Unemployment, total (% of total labor force) (national estimate)
Boxplot for Unemployment, total (% of total labor force) (national estimate)
Density plot for Unemployment, total (% of total labor force) (national estimate)
Frequency
Density
Unemployment, total (% of total labor force) (national estimate)

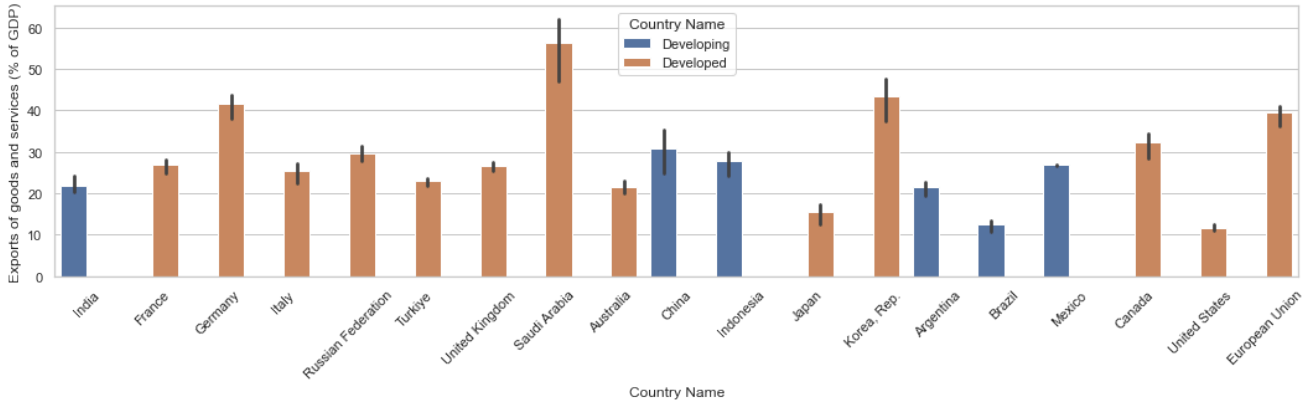

Histogram for Employment in services (% of total employment) (modeled ILO estimate)
Boxplot for Employment in services (% of total employment) (modeled ILO estimate)
Density plot for Employment in services (% of total employment) (modeled ILO estimate)
Frequency
Density
Employment in services (% of total employment) (modeled ILO estimate)

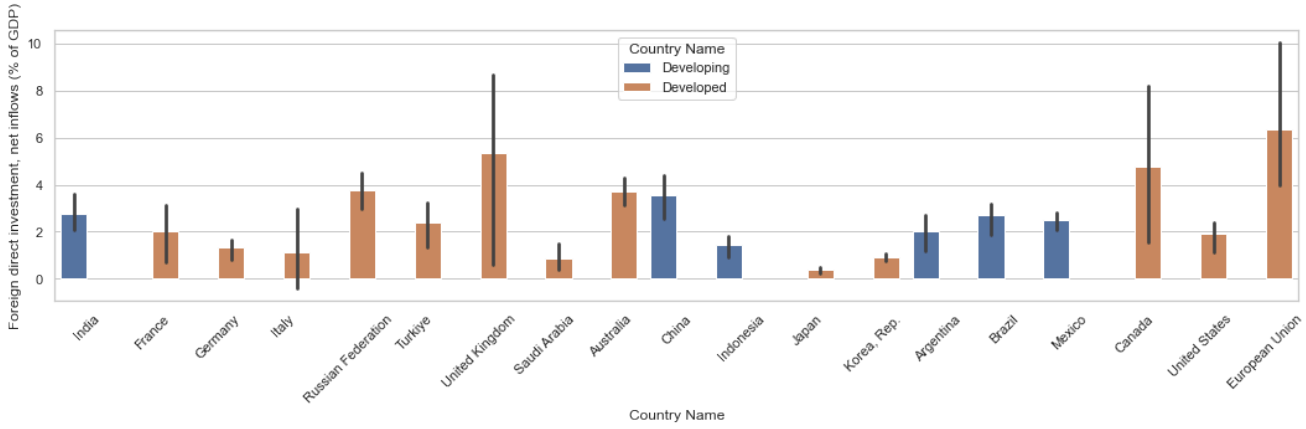

Histogram for Employment in industry (% of total employment) (modeled ILO estimate)
Box plot for Employment in industry (% of total employment) (modeled ILO estimate)
Density plot for Employment in industry (% of total employment) (modeled ILO estimate)
Frequency
Density
Employment in industry (% of total employment) (modeled ILO estimate)

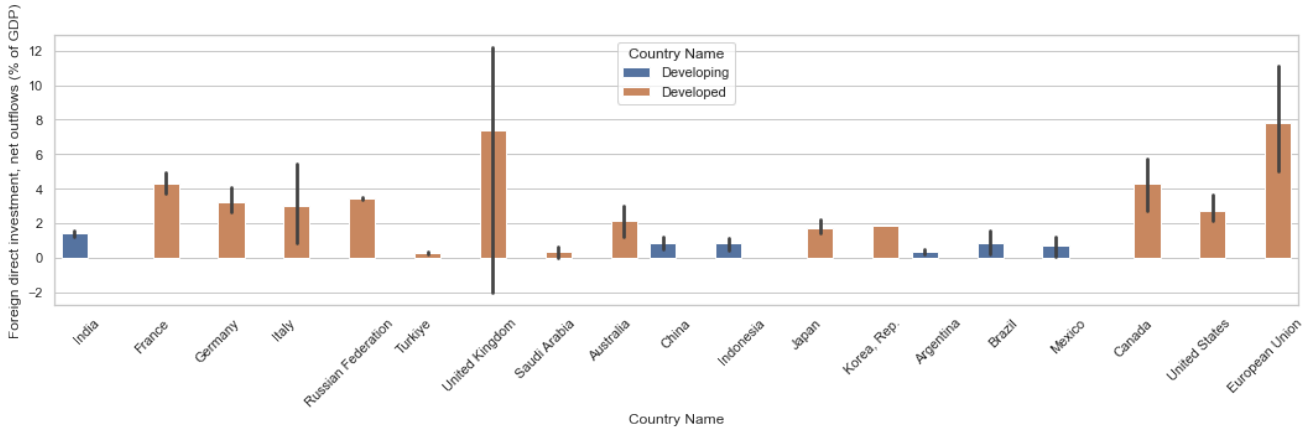

Histogram for Employment in agriculture (% of total employment) (modeled ILO estimate)
Box plot for Employment in agriculture (% of total employment) (modeled ILO estimate)
Density plot for Employment in agriculture (% of total employment) (modeled ILO estimate)
Frequency
Density
Employment in agriculture (% of total employment) (modeled ILO estimate)

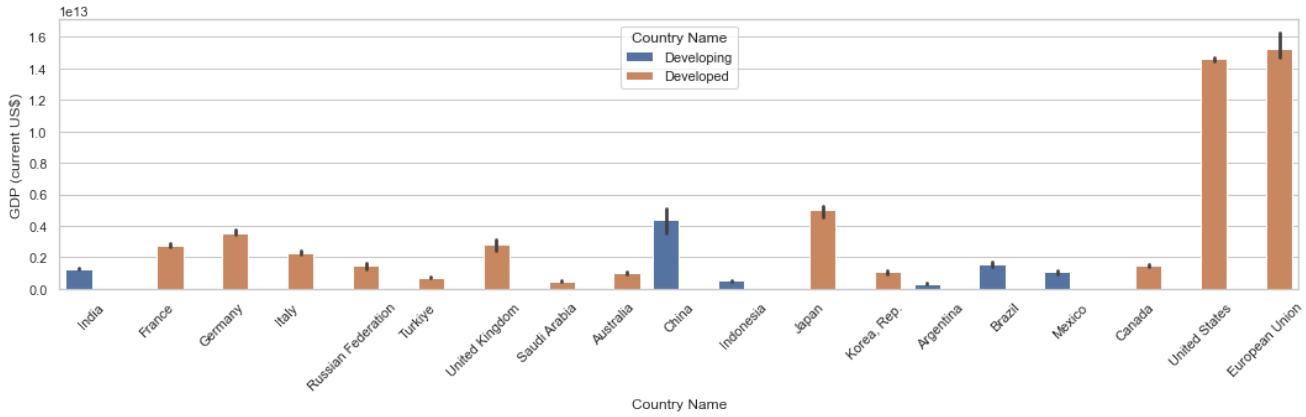
In [89]:
```
developed_countries = ['United States', 'United Kingdom', 'Germany', 'Japan', 'Canada', 'Australia', 'France', 'Italy', 'Russian
developing_countries = ['India', 'Brazil', 'China', 'Mexico', 'South Africa', "Indonesia"]

for i in economic_metrics:
    plt.figure(figsize=(18,4))
    sns.barplot(data=df_economic_crisis, x="Country Name", y=i, hue=df_economic_crisis['Country Name'].apply(lambda x: 'Developed
    plt.xticks(rotation=45)
    plt.show()
```

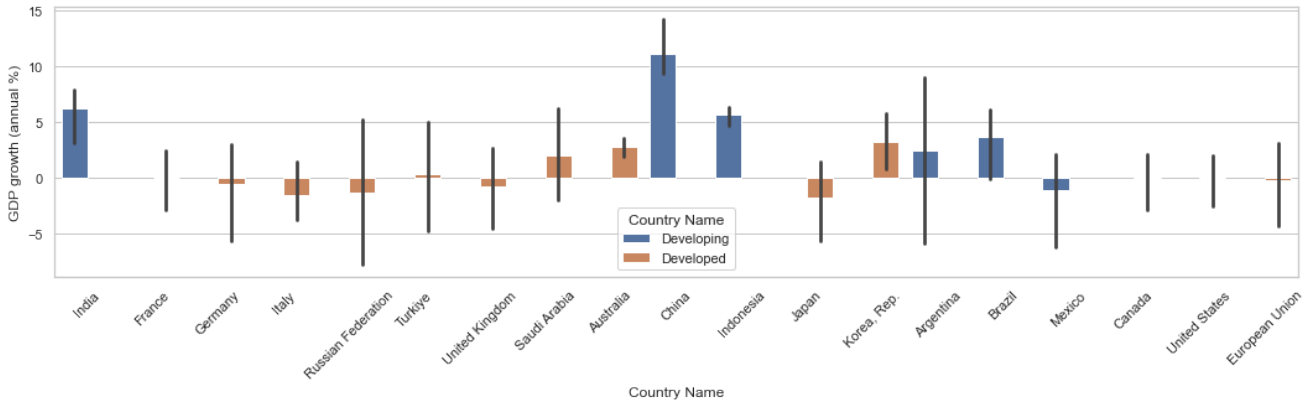


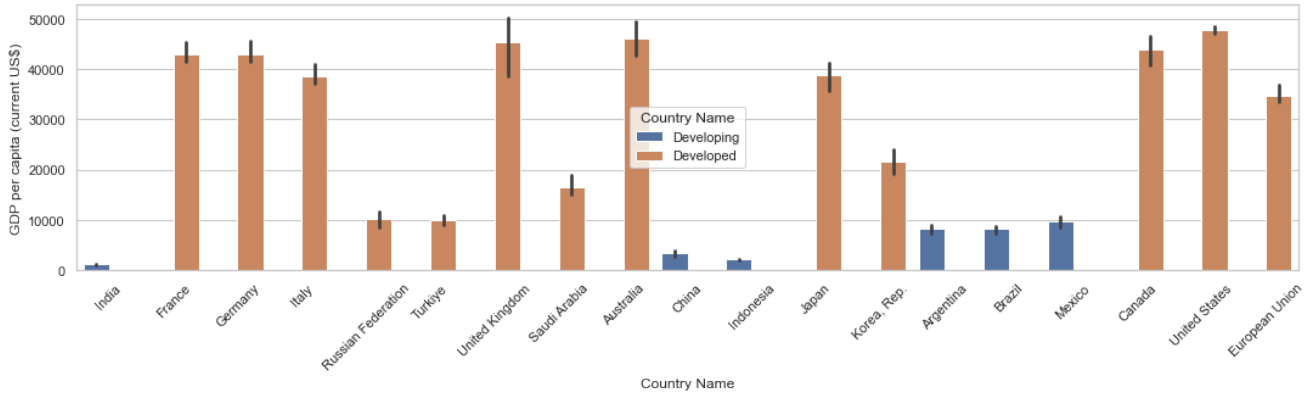


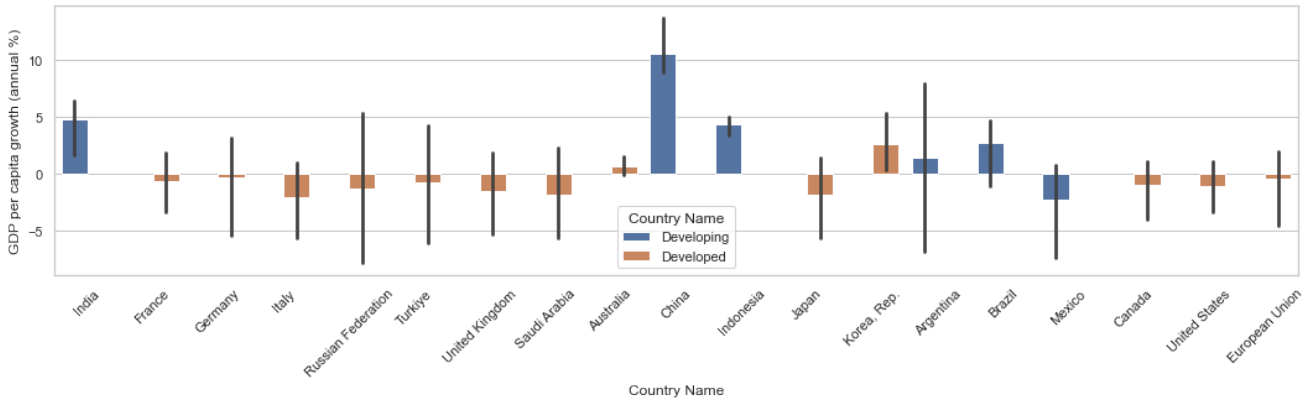


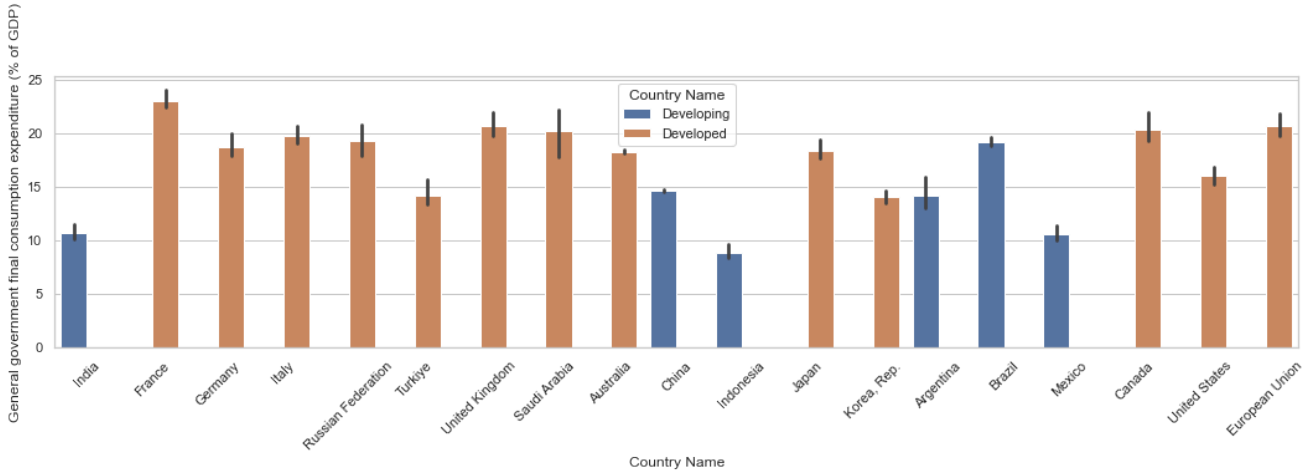

Foreign direct investment, net outflows (% of GDP)
Country Name
Developing
Developed
India
France
Germany
Italy
Russian Federation
Turkiye
United Kingdom
Saudi Arabia
Australia
China
Indonesia
Japan
Korea, Rep.
Argentina
Brazil
Mexico
Canada
United States
European Union
Country Name

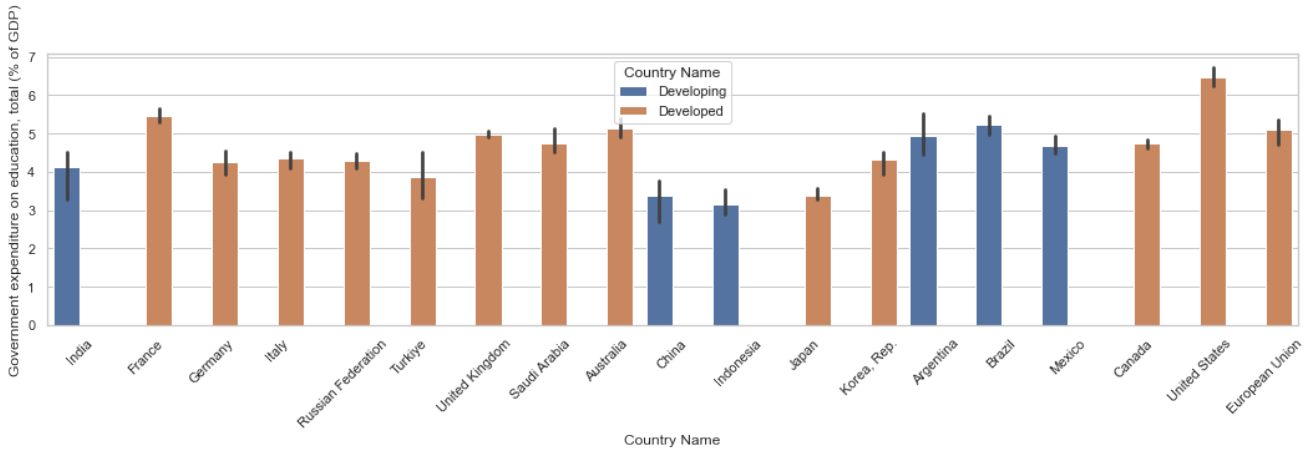

1e13
GDP (current US$)
Country Name
Developing
Developed
India
France
Germany
Italy
Russian Federation
Turkiye
United Kingdom
Saudi Arabia
Australia
China
Indonesia
Japan
Korea, Rep.
Argentina
Brazil
Mexico
Canada
United States
European Union
Country Name

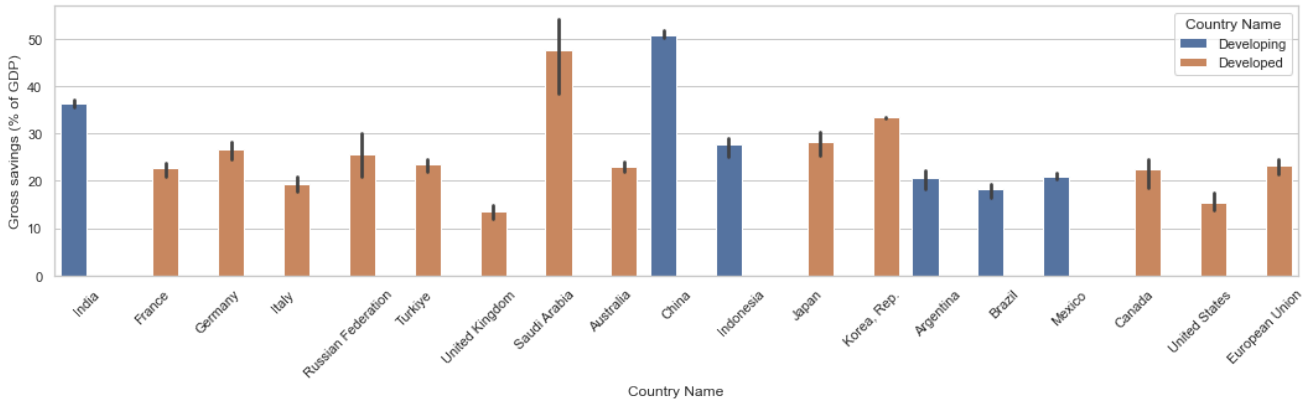

GDP growth (annual %)
Country Name
Developing
Developed
India
France
Germany
Italy
Russian Federation
Turkiye
United Kingdom
Saudi Arabia
Australia
China
Indonesia
Japan
Korea, Rep.
Argentina
Brazil
Mexico
Canada
United States
European Union
Country Name

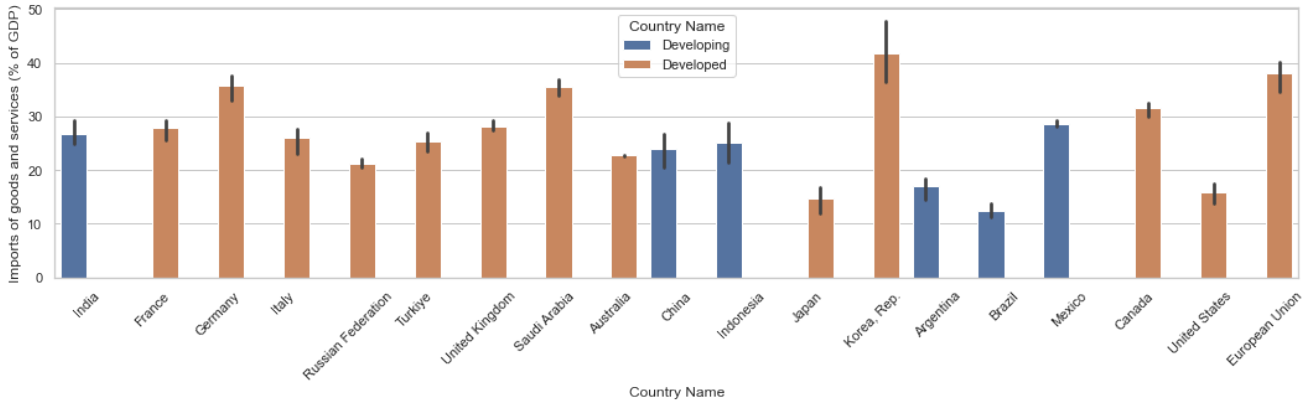

GDP per capita (current US$)
Country Name
Developing
Developed
India
France
Germany
Italy
Russian Federation
Turkiye
United Kingdom
Saudi Arabia
Australia
China
Indonesia
Japan
Korea, Rep.
Argentina
Brazil
Mexico
Canada
United States
European Union
Country Name

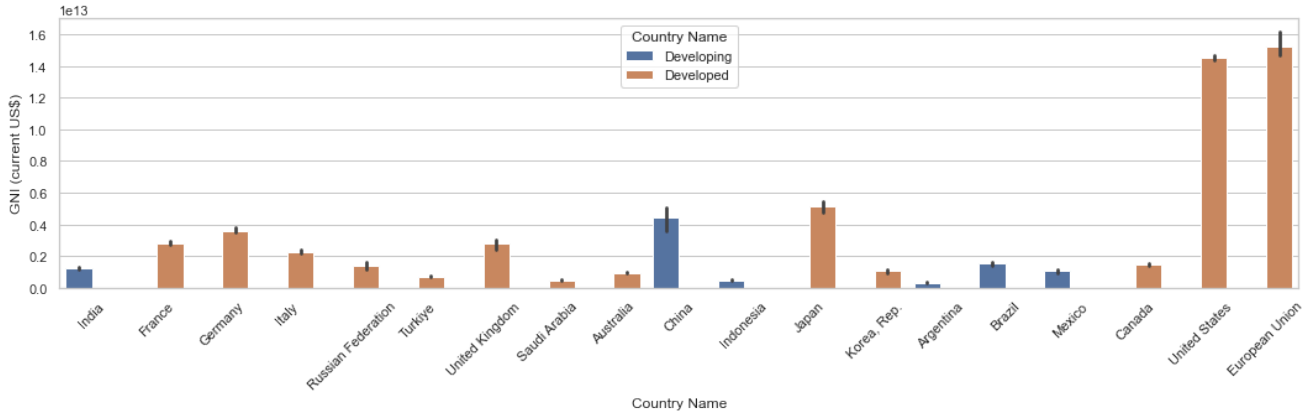

GDP per capita growth (annual %)
Country Name
Developing
Developed
India
France
Germany
Italy
Russian Federation
Turkiye
United Kingdom
Saudi Arabia
Australia
China
Indonesia
Japan
Korea, Rep.
Argentina
Brazil
Mexico
Canada
United States
European Union
Country Name

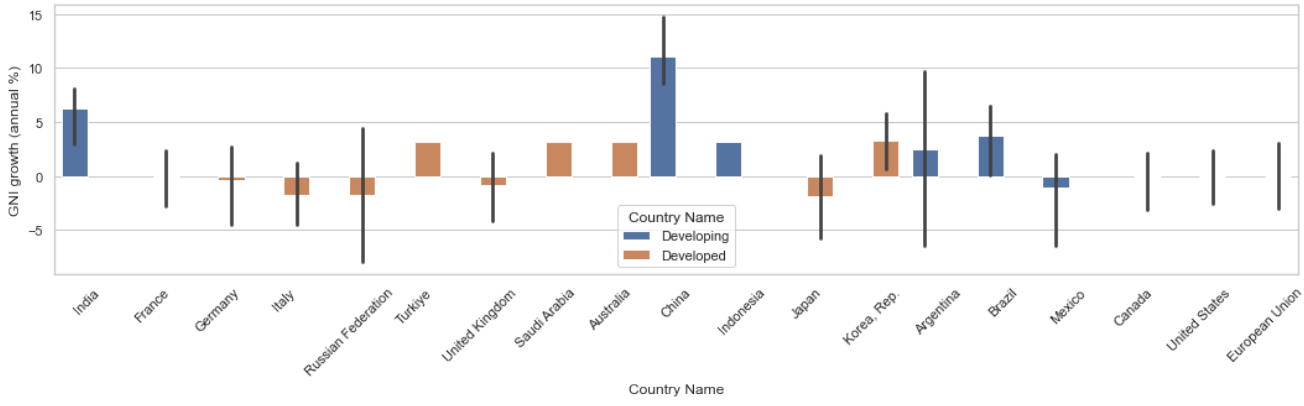

General government final consumption expenditure (% of GDP)
Country Name
Developing
Developed
India
France
Germany
Italy
Russian Federation
Turkiye
United Kingdom
Saudi Arabia
Australia
China
Indonesia
Japan
Korea, Rep.
Argentina
Brazil
Mexico
Canada
United States
European Union
Country Name

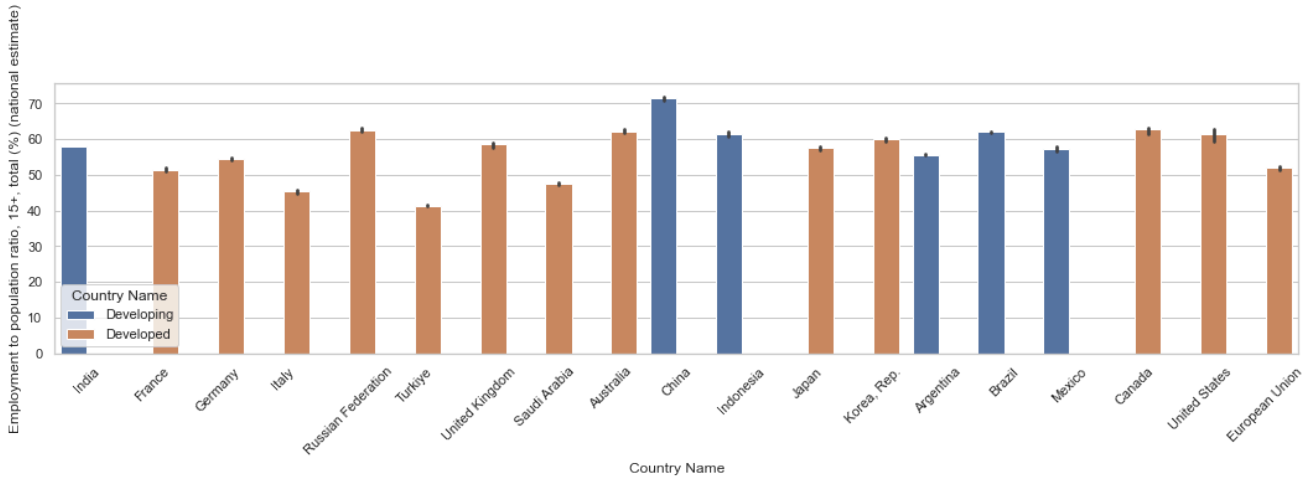

Government expenditure on education, total (% of GDP)
Country Name
Developing
Developed
India
France
Germany
Italy
Russian Federation
Turkiye
United Kingdom
Saudi Arabia
Australia
China
Indonesia
Japan
Korea, Rep.
Argentina
Brazil
Mexico
Canada
United States
European Union
Country Name

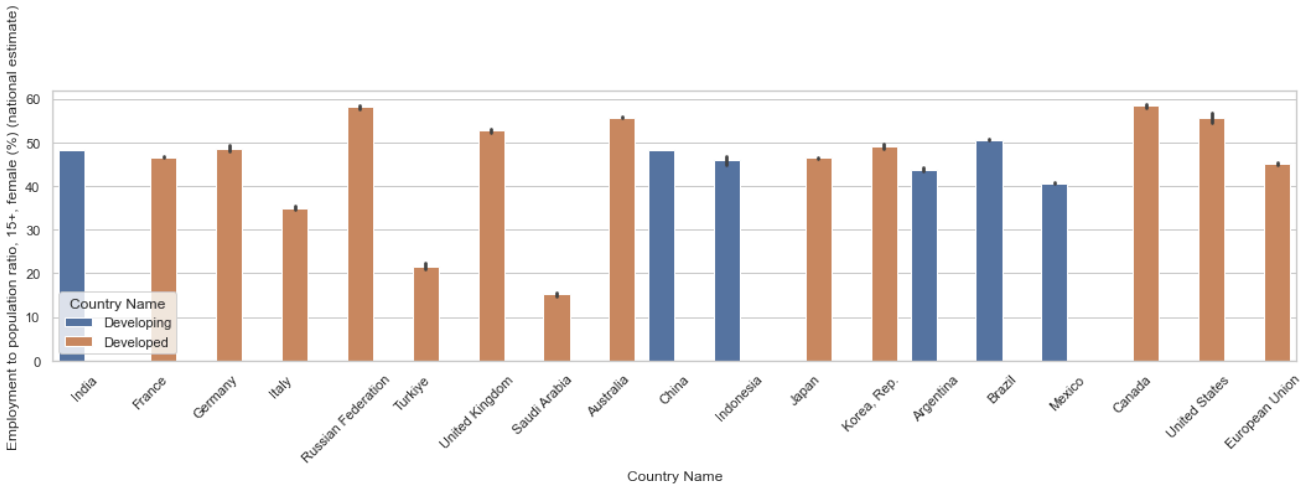

Gross savings (% of GDP)
Country Name
Developing
Developed
India
France
Germany
Italy
Russian Federation
Turkiye
United Kingdom
Saudi Arabia
Australia
China
Indonesia
Japan
Korea, Rep.
Argentina
Brazil
Mexico
Canada
United States
European Union
Country Name

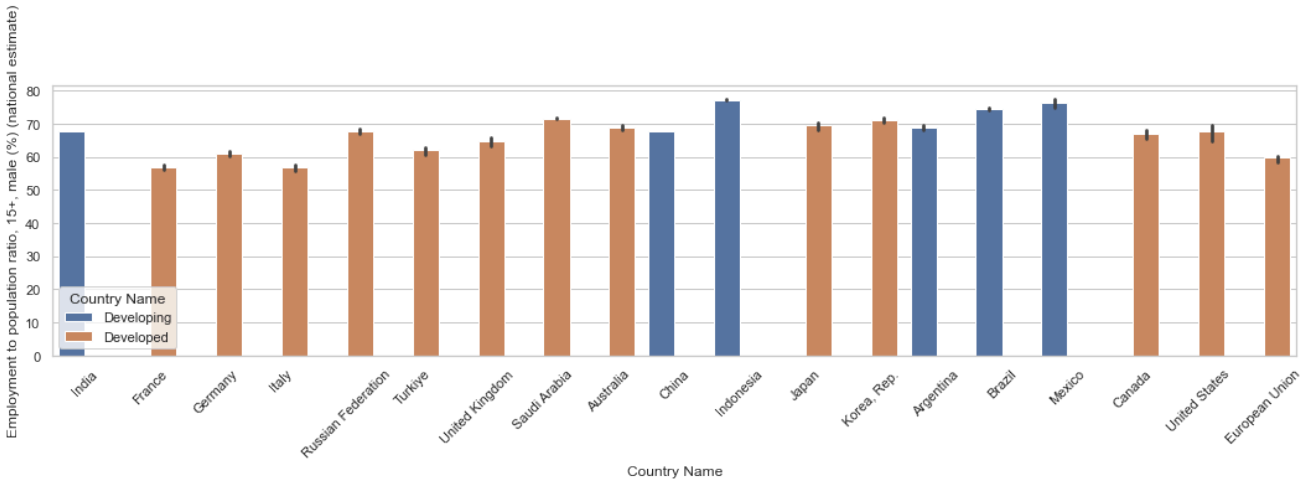

Country Name
Developing
Developed
Imports of goods and services (% of GDP)
Country Name
India
France
Germany
Italy
Russian Federation
Turkiye
United Kingdom
Saudi Arabia
Australia
China
Indonesia
Japan
Korea, Rep.
Argentina
Brazil
Mexico
Canada
United States
European Union

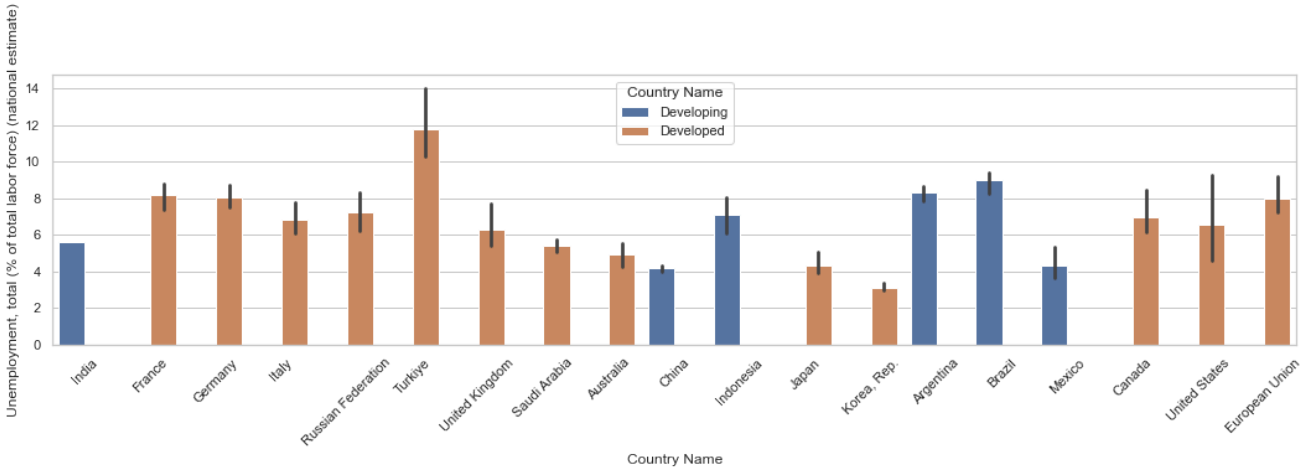

1e13
Country Name
Developing
Developed
GNI (current US$)
Country Name
India
France
Germany
Italy
Russian Federation
Turkiye
United Kingdom
Saudi Arabia
Australia
China
Indonesia
Japan
Korea, Rep.
Argentina
Brazil
Mexico
Canada
United States
European Union

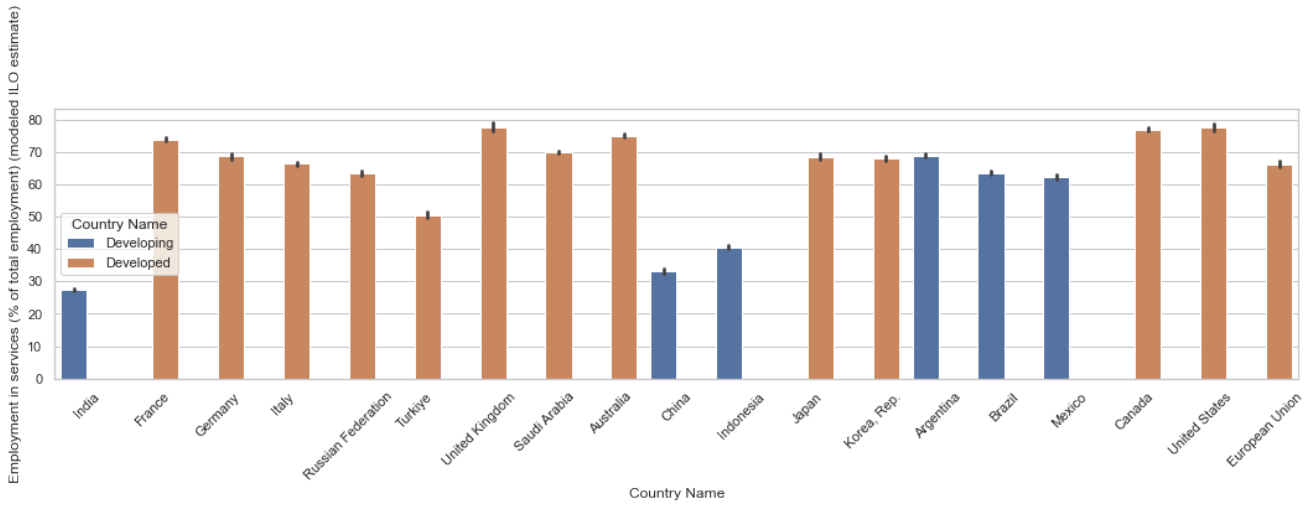

Country Name
Developing
Developed
GNI growth (annual %)
Country Name
India
France
Germany
Italy
Russian Federation
Turkiye
United Kingdom
Saudi Arabia
Australia
China
Indonesia
Japan
Korea, Rep.
Argentina
Brazil
Mexico
Canada
United States
European Union

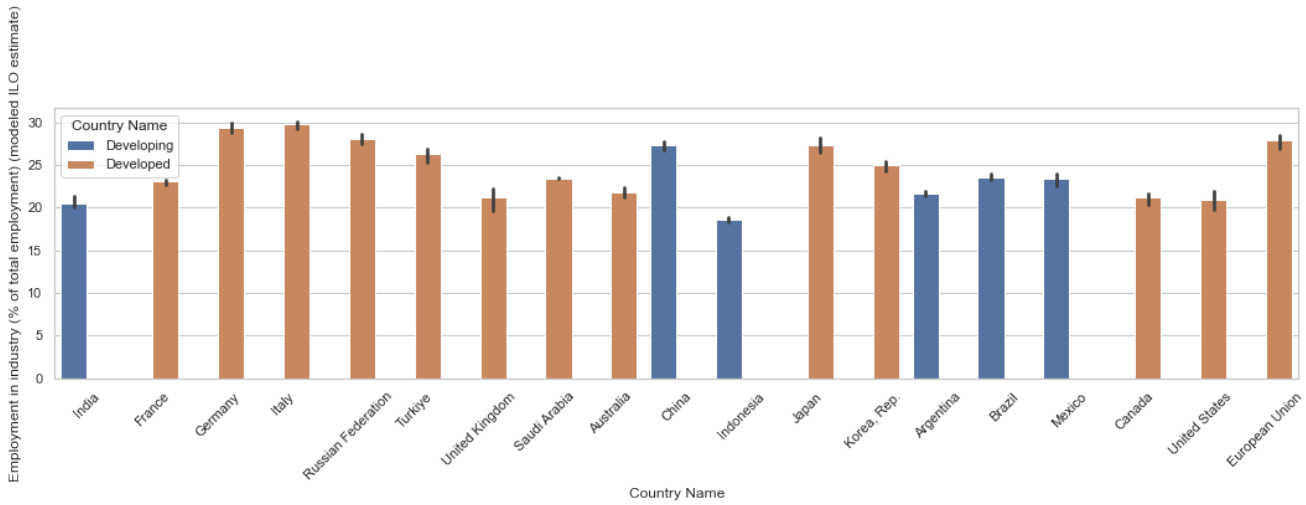

Country Name
Developing
Developed
Employment to population ratio, 15+, total (%) (national estimate)
Country Name
India
France
Germany
Italy
Russian Federation
Turkiye
United Kingdom
Saudi Arabia
Australia
China
Indonesia
Japan
Korea, Rep.
Argentina
Brazil
Mexico
Canada
United States
European Union

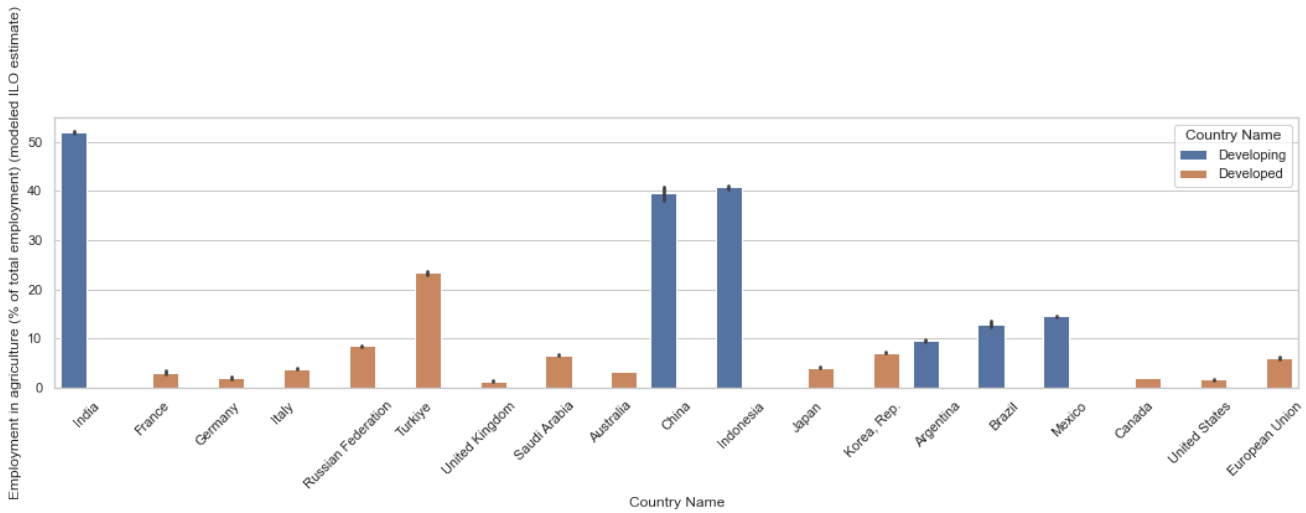

Employment to population ratio, 15+, female (%) (national estimate)
Country Name
Developing
Developed
India
France
Germany
Italy
Russian Federation
Turkiye
United Kingdom
Saudi Arabia
Australia
China
Indonesia
Japan
Korea, Rep.
Argentina
Brazil
Mexico
Canada
United States
European Union
Country Name


Employment to population ratio, 15+, male (%) (national estimate)
Country Name
Developing
Developed
India
France
Germany
Italy
Russian Federation
Turkiye
United Kingdom
Saudi Arabia
Australia
China
Indonesia
Japan
Korea, Rep.
Argentina
Brazil
Mexico
Canada
United States
European Union
Country Name

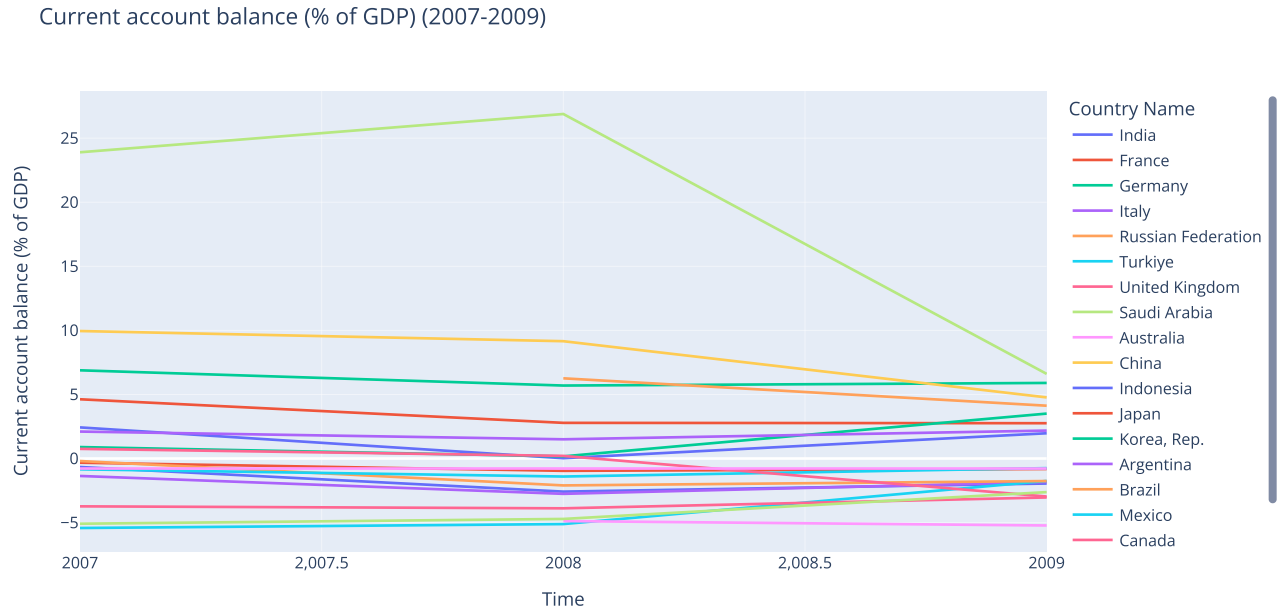

Unemployment, total (% of total labor force) (national estimate)
Country Name
Developing
Developed
India
France
Germany
Italy
Russian Federation
Turkiye
United Kingdom
Saudi Arabia
Australia
China
Indonesia
Japan
Korea, Rep.
Argentina
Brazil
Mexico
Canada
United States
European Union
Country Name


Employment in services (% of total employment) (modeled ILO estimate)
Country Name
Developing
Developed
India
France
Germany
Italy
Russian Federation
Turkiye
United Kingdom
Saudi Arabia
Australia
China
Indonesia
Japan
Korea, Rep.
Argentina
Brazil
Mexico
Canada
United States
European Union
Country Name

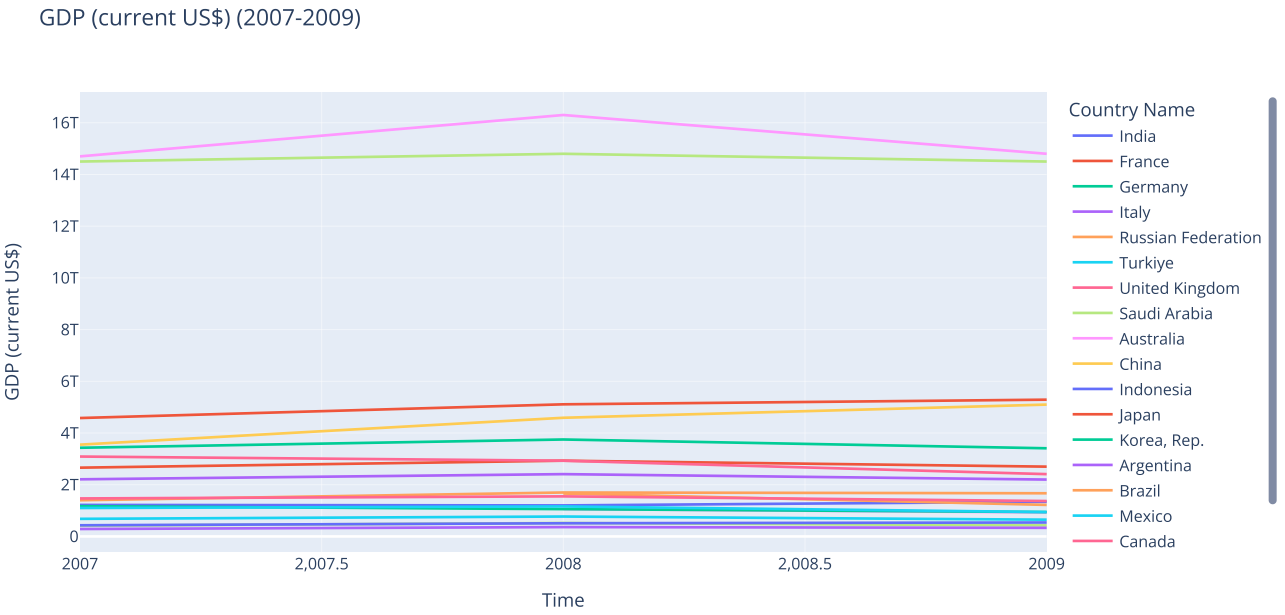

Employment in industry (% of total employment) (modeled ILO estimate)
Country Name
Developing
Developed
India
France
Germany
Italy
Russian Federation
Turkiye
United Kingdom
Saudi Arabia
Australia
China
Indonesia
Japan
Korea, Rep.
Argentina
Brazil
Mexico
Canada
United States
European Union
Country Name


Employment in agriculture (% of total employment) (modeled ILO estimate)
Country Name
Developing
Developed
India
France
Germany
Italy
Russian Federation
Turkiye
United Kingdom
Saudi Arabia
Australia
China
Indonesia
Japan
Korea, Rep.
Argentina
Brazil
Mexico
Canada
United States
European Union
Country Name

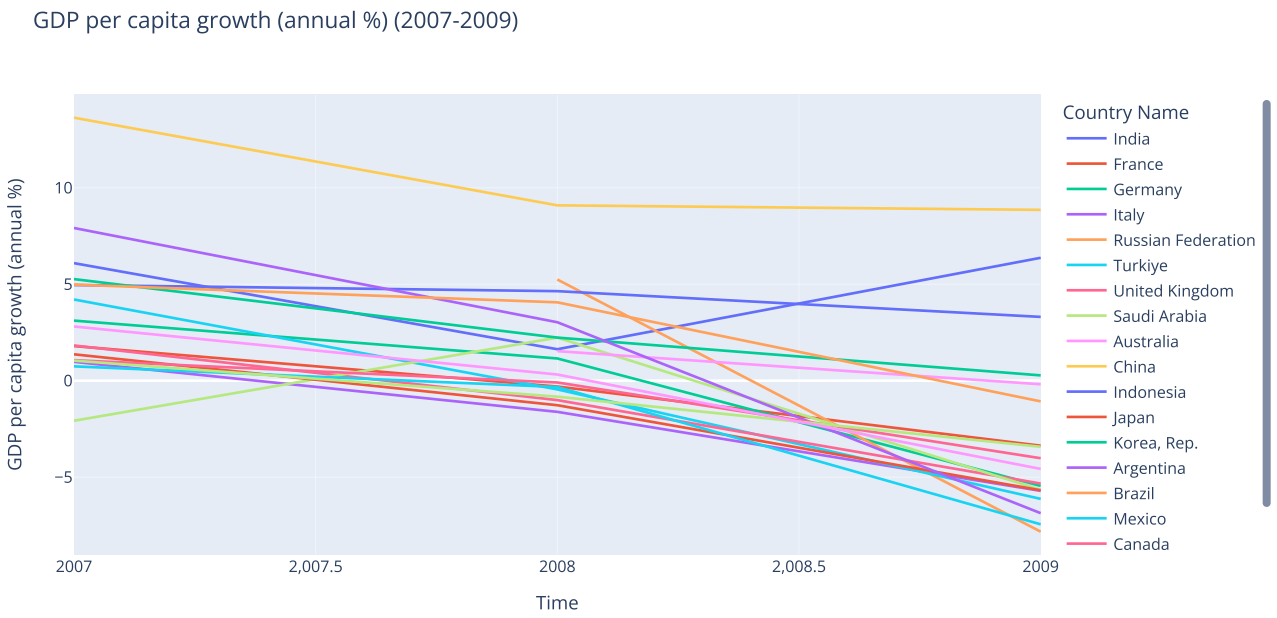
In [95]:
```
fig = px.line(df_economic_crisis, x='Time', y='Current account balance (% of GDP)', color='Country Name',
              title=f"Current account balance (% of GDP) (2007-2009)")
fig.show()
```



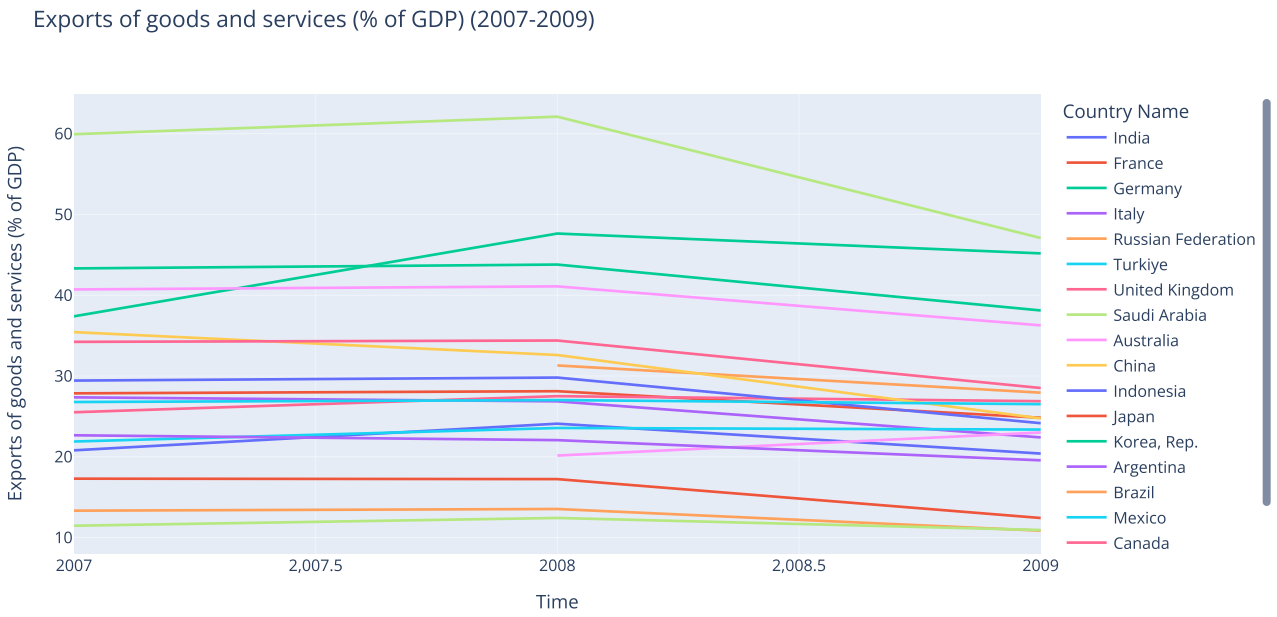
In [96]:
```
fig = px.line(df_economic_crisis, x='Time', y='GDP (current US$)', color='Country Name',
              title=f"GDP (current US$) (2007-2009)")
fig.show()
```



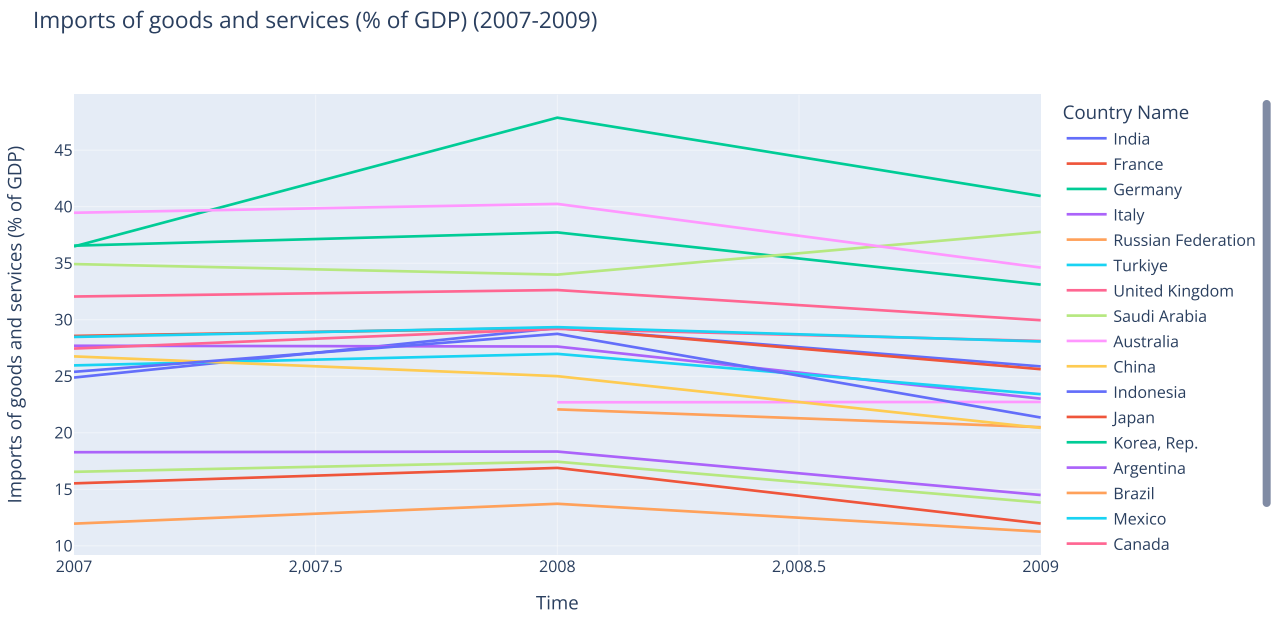
In [97]:
```
fig = px.line(df_economic_crisis, x='Time', y='GDP per capita growth (annual %)', color='Country Name',
              title=f"GDP per capita growth (annual %) (2007-2009)")
fig.show()
```



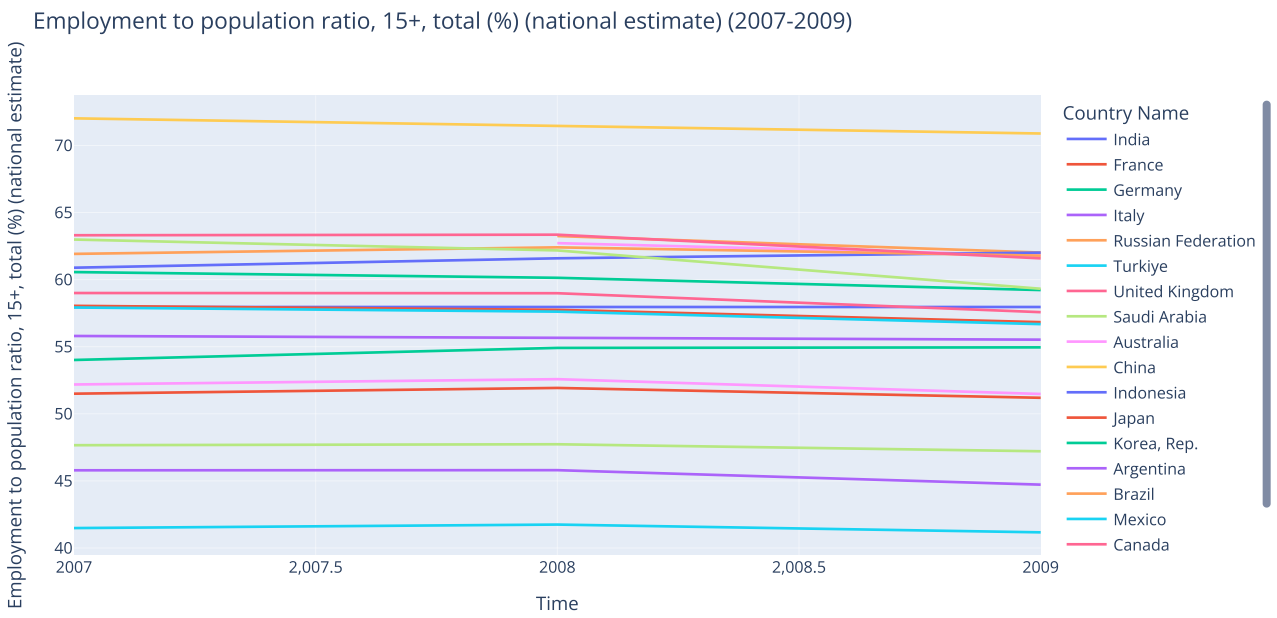
In [98]:
```
fig = px.line(df_economic_crisis, x='Time', y='Exports of goods and services (% of GDP)', color='Country Name',
              title=f"Exports of goods and services (% of GDP) (2007-2009)")
fig.show()
```



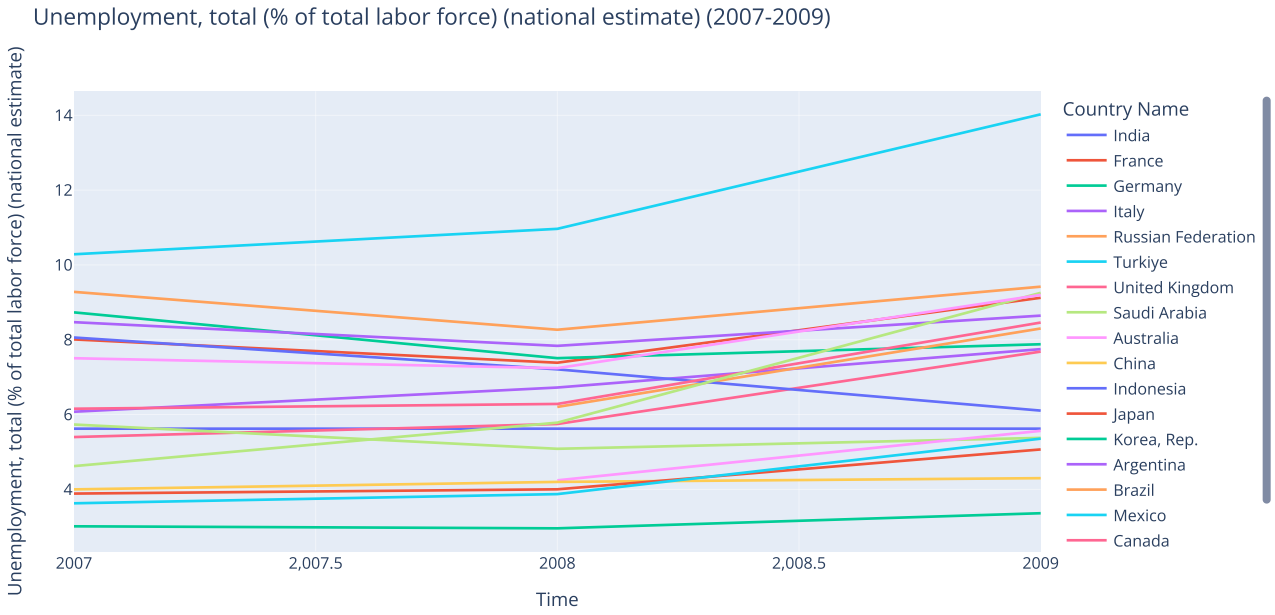
In [99]:
```
fig = px.line(df_economic_crisis, x='Time', y='Imports of goods and services (% of GDP)', color='Country Name',
              title=f"Imports of goods and services (% of GDP) (2007-2009)")
fig.show()
```



In [100]:
```
fig = px.line(df_economic_crisis, x='Time', y='Employment to population ratio, 15+, total (%) (national estimate)', color='Countr
              title=f"Employment to population ratio, 15+, total (%) (national estimate) (2007-2009)")
fig.show()
```



In [101]:
```
fig = px.line(df_economic_crisis, x='Time', y='Unemployment, total (% of total labor force) (national estimate)', color='Country
              title=f"Unemployment, total (% of total labor force) (national estimate) (2007-2009)")
fig.show()
```



Case Study 2: COVID-19 Pandemic

The COVID-19 pandemic (also known as the coronavirus pandemic and COVID pandemic), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), began with an outbreak of COVID-19 in Wuhan, China, in December 2019. It spread to other areas of Asia, and then worldwide in early 2020. The World Health Organization (WHO) declared the outbreak a public health emergency of international concern (PHEIC) on 30 January 2020, and assessed the outbreak as having become a pandemic on 11 March.

The pandemic and responses to it damaged the global economy. On 27 February 2020, worries about the outbreak crushed US stock indexes, which posted their sharpest falls since 2008.

Tourism collapsed due to travel restrictions, closing of public places including travel attractions, and advice of governments against travel. Airlines cancelled flights, while British regional airline Flybe collapsed. The cruise line industry was hard hit, and train stations and ferry ports closed. International mail stopped or was delayed

We will conduct a covid case study on G20 nations

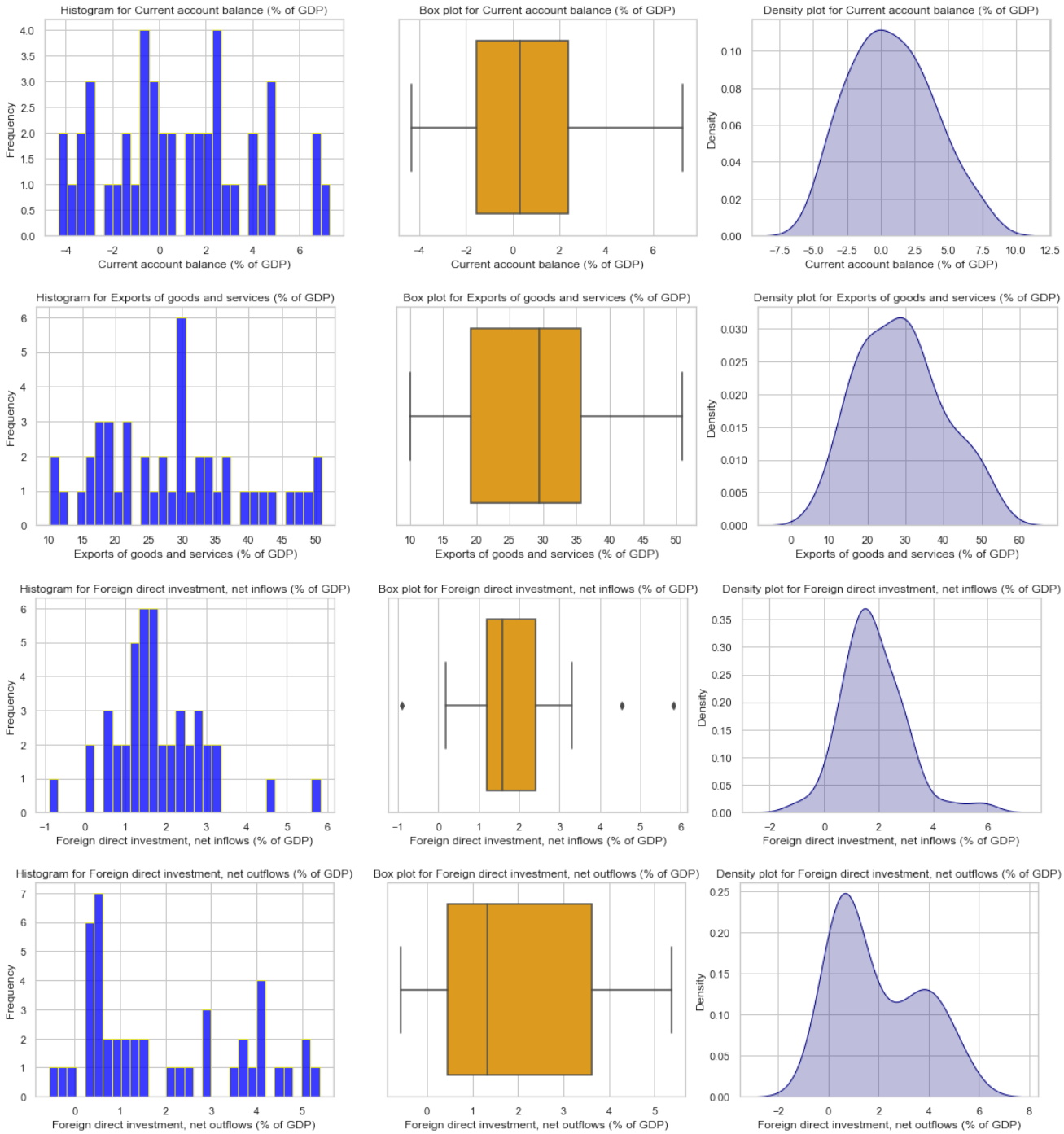
In [104]:
```
g20 = ['Argentina', 'Australia', 'Brazil', 'Canada', 'China', 'France', 'Germany', 'India', 'Indonesia', 'Italy', 'Japan', 'Mexic
for i in g20:
    if i not in df["Country Name"].unique():
        print(i)

years = [2020, 2021, 2022]

df_covid_crisis = df1[(df1["Country Name"].isin(g20)) & (df1["Time"].isin(years))]
df_covid_crisis.head()
```

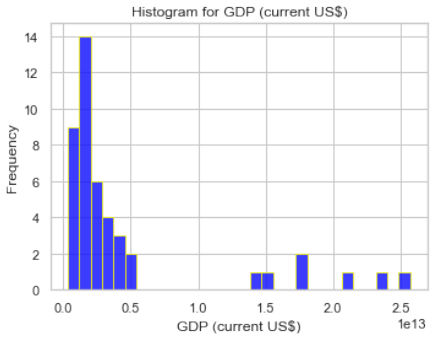
Out[104]:

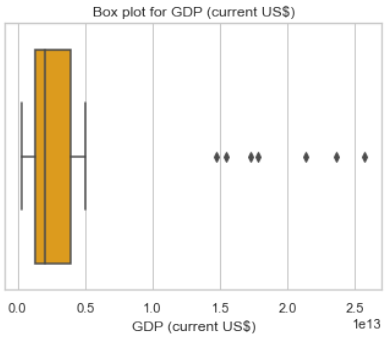
	Country Name	Country Code	Time	Current account balance (% of GDP)	Exports of goods and services (% of GDP)	Foreign direct investment, net inflows (% of GDP)	Foreign direct investment, net outflows (% of GDP)	GDP (current US$)	GDP growth (annual %)	GDP per capita (current US$)	...	Current health expenditure per capita (current US$)	Hospital beds (per 1,000 people)	Physicians (per 1,000 people)	exp
66	India	IND	2020	1.223621	18.682477	2.406203	0.415824	2.670000e+12	-5.777725	1915.551588	...	64.000000	1.620	0.727	7(
67	India	IND	2021	-1.055242	21.399158	1.412171	0.544277	3.170000e+12	9.689592	2250.179018	...	74.000000	1.600	2.289	6
204	France	FRA	2020	-1.563074	27.327736	0.731612	1.085462	2.650000e+12	-7.540459	39179.744260	...	4755.440430	6.000	3.324	8
205	France	FRA	2021	0.333456	30.034994	3.311955	3.994612	2.960000e+12	6.435210	43671.308410	...	5380.875000	2.795	2.289	8
227	Germany	DEU	2020	6.567299	43.476980	4.549376	3.729155	3.890000e+12	-3.826769	46749.476230	...	5936.251465	7.800	4.459	8

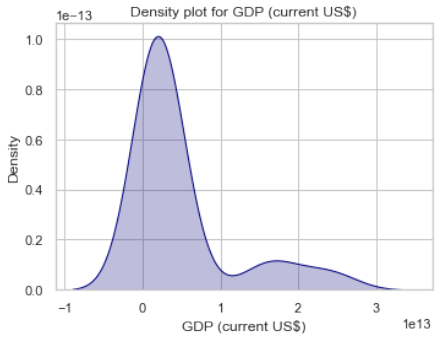
5 rows × 48 columns

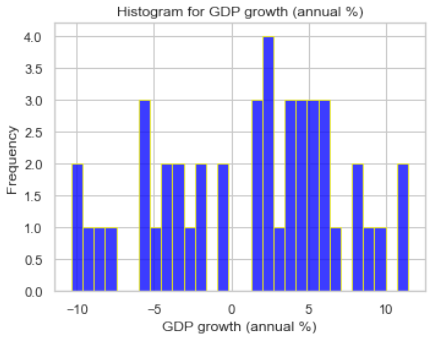
In [105]:
```
for i in economic_metrics:
    plt.figure(figsize=(15,4))

    plt.subplot(1,3,1)
    sns.histplot(data=df_covid_crisis[i], color="blue", edgecolor="yellow", bins=30)
    plt.xlabel(i)
    plt.ylabel("Frequency")
    plt.title(f"Histogram for {i}")

    plt.subplot(1,3,2)
    sns.boxplot(x=df_covid_crisis[i], color="orange")
    plt.xlabel(i)
    plt.title(f"Box plot for {i}")

    plt.subplot(1,3,3)
    sns.kdeplot(data=df_covid_crisis[i], color="navy", fill=True)
    plt.xlabel(i)
    plt.ylabel("Density")
    plt.title(f"Density plot for {i}")

    plt.tight_layout()
    plt.show()
```

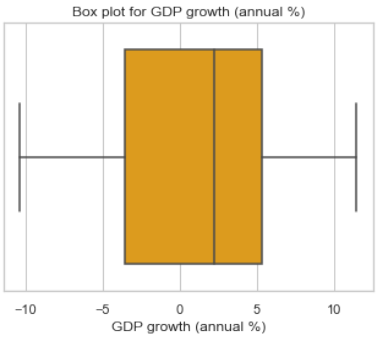


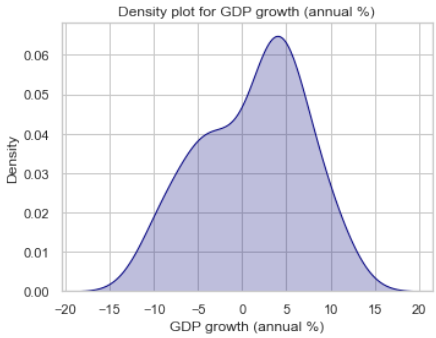

Histogram for GDP (current US$)
Frequency
GDP (current US$)
1e13

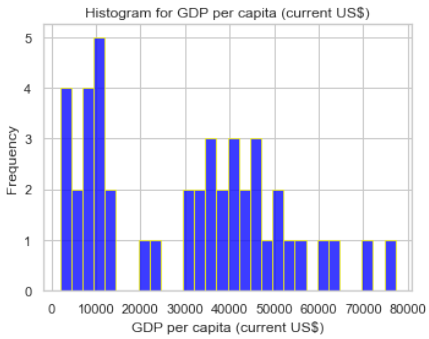

Box plot for GDP (current US$)
GDP (current US$)
1e13

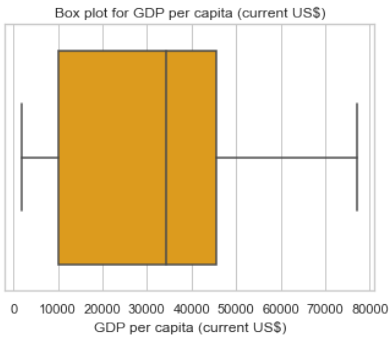

1e−13
Density plot for GDP (current US$)
Density
GDP (current US$)
1e13

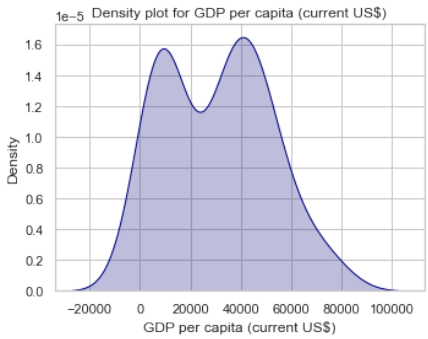

Histogram for GDP growth (annual %)
Frequency
GDP growth (annual %)

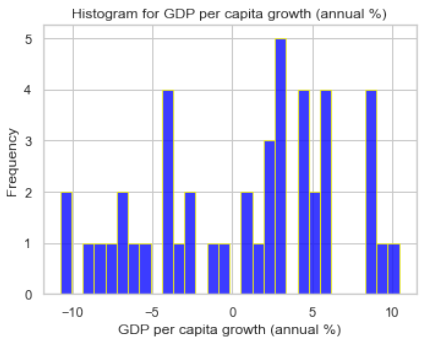

Box plot for GDP growth (annual %)
GDP growth (annual %)

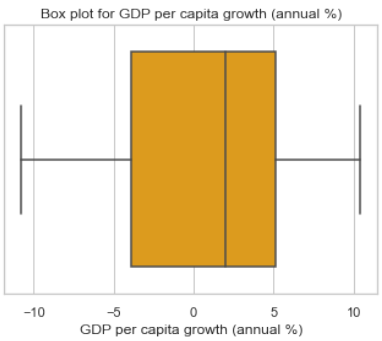

Density plot for GDP growth (annual %)
Density
GDP growth (annual %)

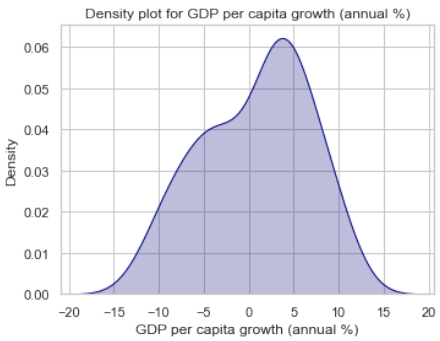

Histogram for GDP per capita (current US$)
Frequency
GDP per capita (current US$)

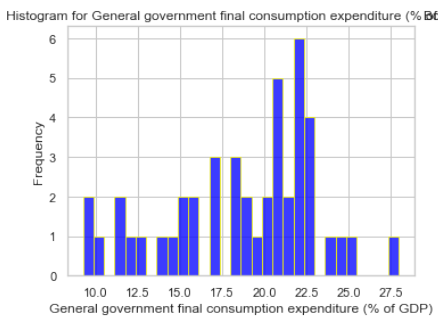

Box plot for GDP per capita (current US$)
GDP per capita (current US$)

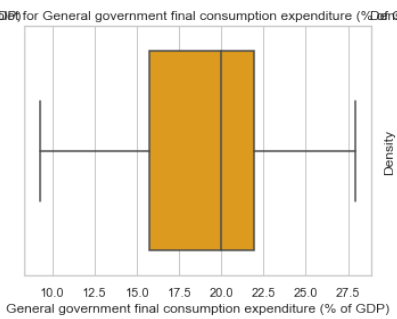

1e−5
Density plot for GDP per capita (current US$)
Density
GDP per capita (current US$)

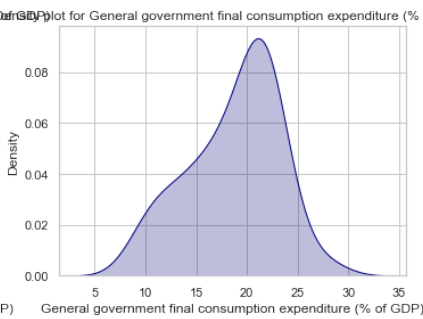

Histogram for GDP per capita growth (annual %)
Frequency
GDP per capita growth (annual %)

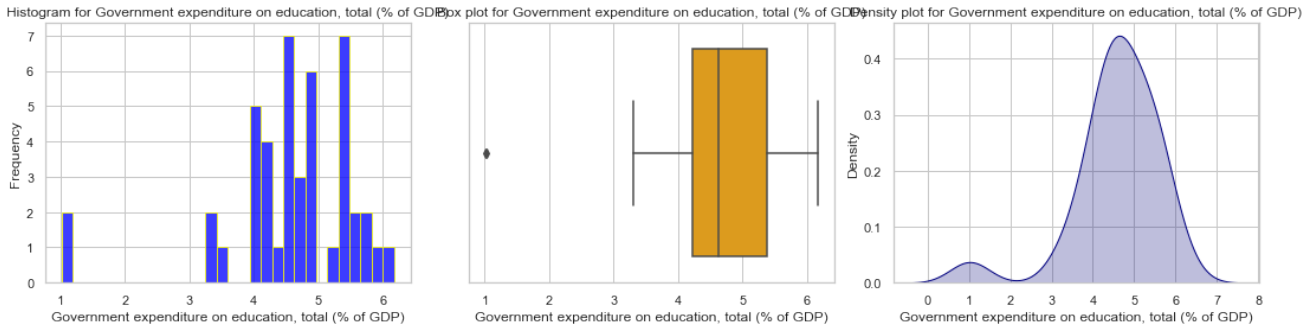

Box plot for GDP per capita growth (annual %)
GDP per capita growth (annual %)

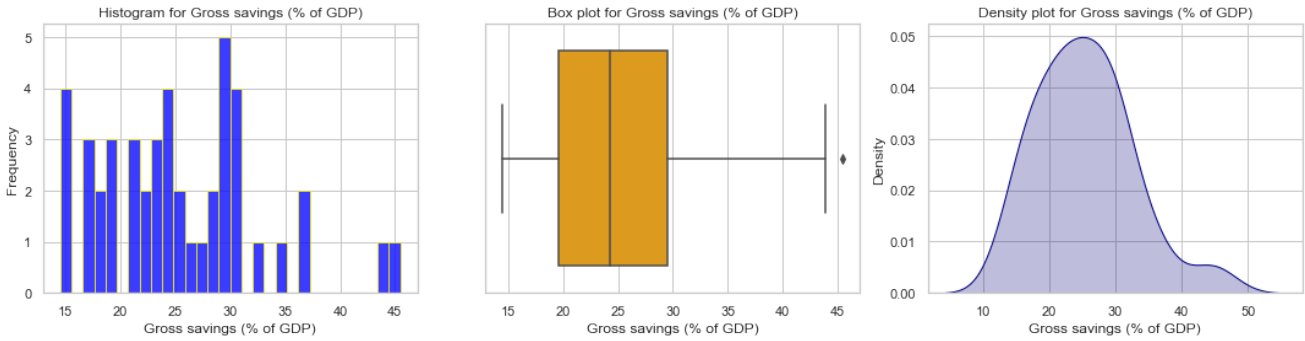

Density plot for GDP per capita growth (annual %)
Density
GDP per capita growth (annual %)

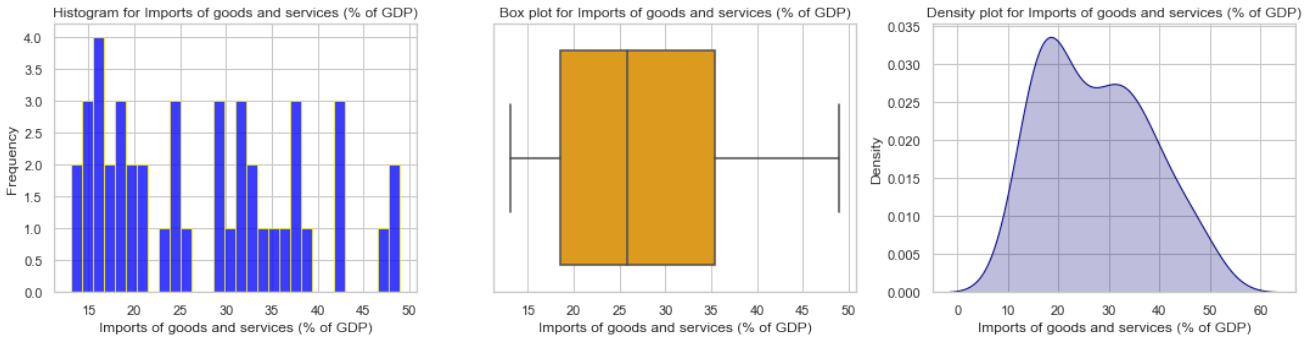

Histogram for General government final consumption expenditure (% of GDP)
Frequency
General government final consumption expenditure (% of GDP)

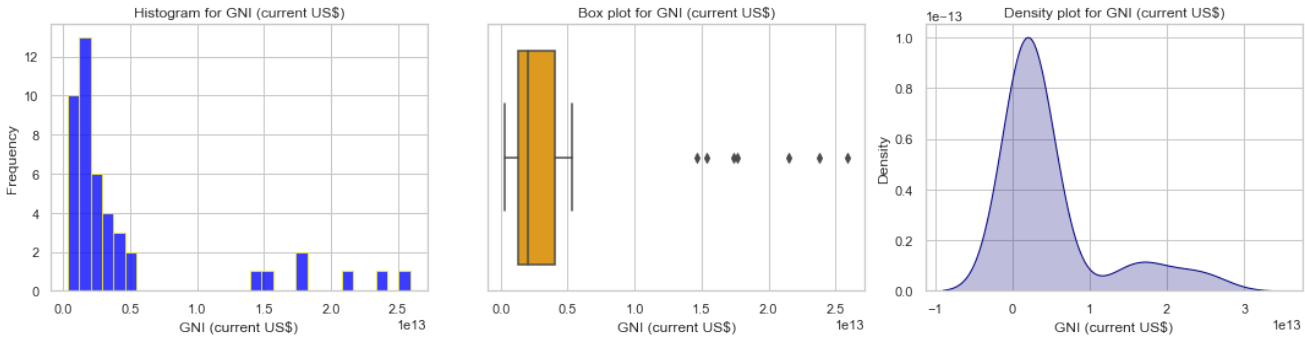

Box plot for General government final consumption expenditure (% of GDP)
General government final consumption expenditure (% of GDP)

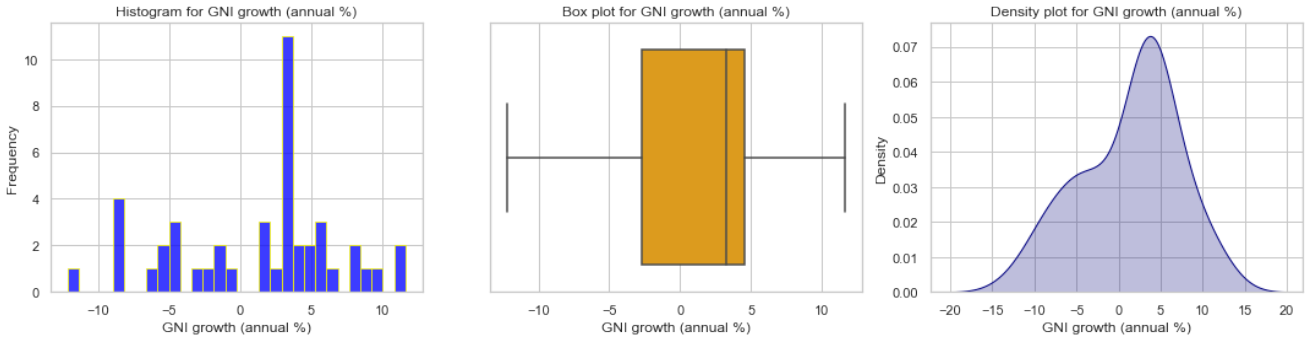

Density plot for General government final consumption expenditure (% of GDP)
Density
General government final consumption expenditure (% of GDP)

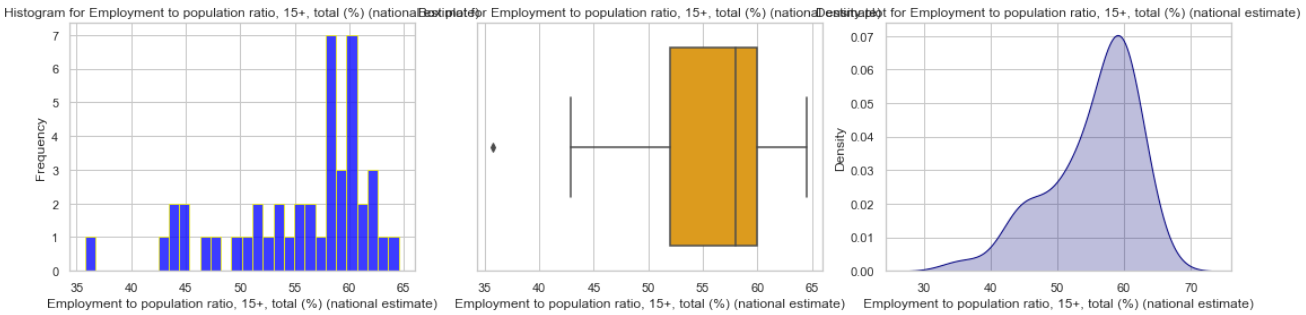

Histogram for Government expenditure on education, total (% of GDP)
Frequency
Government expenditure on education, total (% of GDP)
Box plot for Government expenditure on education, total (% of GDP)
Government expenditure on education, total (% of GDP)
Density plot for Government expenditure on education, total (% of GDP)
Density
Government expenditure on education, total (% of GDP)

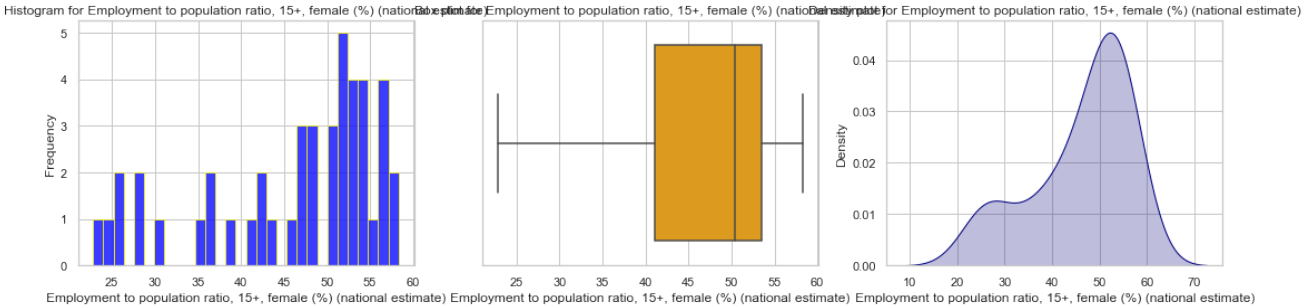

Histogram for Gross savings (% of GDP)
Frequency
Gross savings (% of GDP)
Box plot for Gross savings (% of GDP)
Gross savings (% of GDP)
Density plot for Gross savings (% of GDP)
Density
Gross savings (% of GDP)

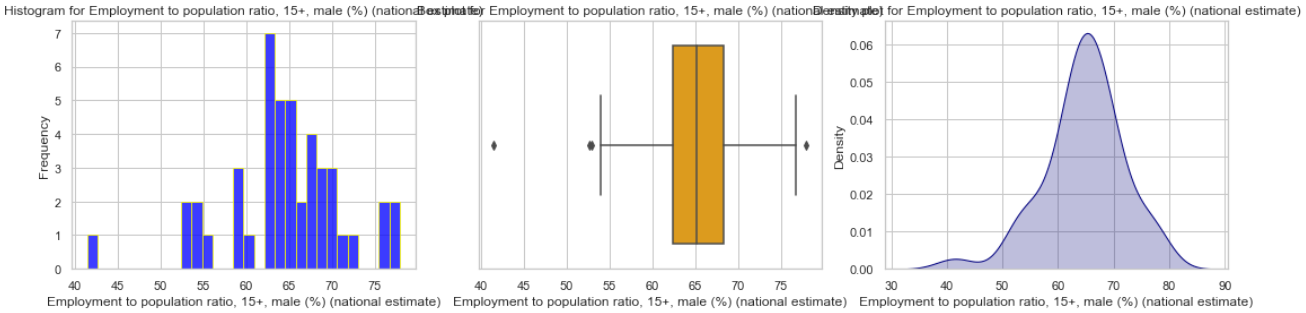

Histogram for Imports of goods and services (% of GDP)
Frequency
Imports of goods and services (% of GDP)
Box plot for Imports of goods and services (% of GDP)
Imports of goods and services (% of GDP)
Density plot for Imports of goods and services (% of GDP)
Density
Imports of goods and services (% of GDP)

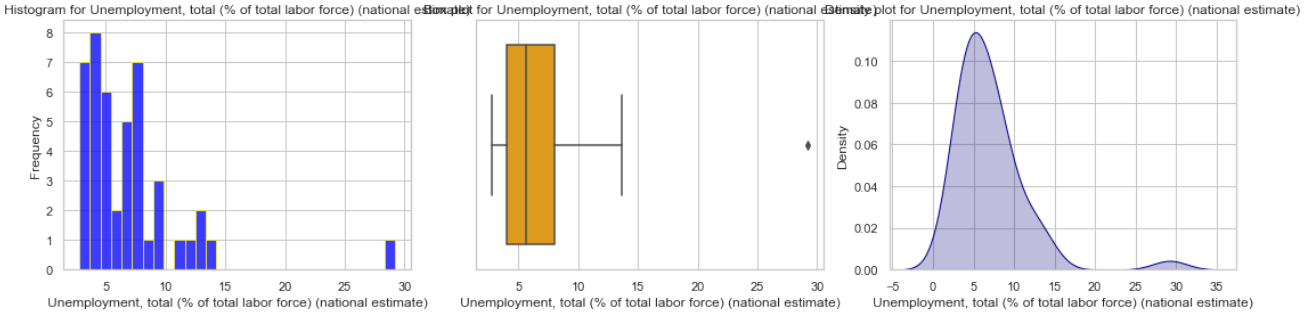

Histogram for GNI (current US$)
Frequency
GNI (current US$)
1e13
Box plot for GNI (current US$)
GNI (current US$)
1e13
Density plot for GNI (current US$)
1e−13
Density
GNI (current US$)
1e13

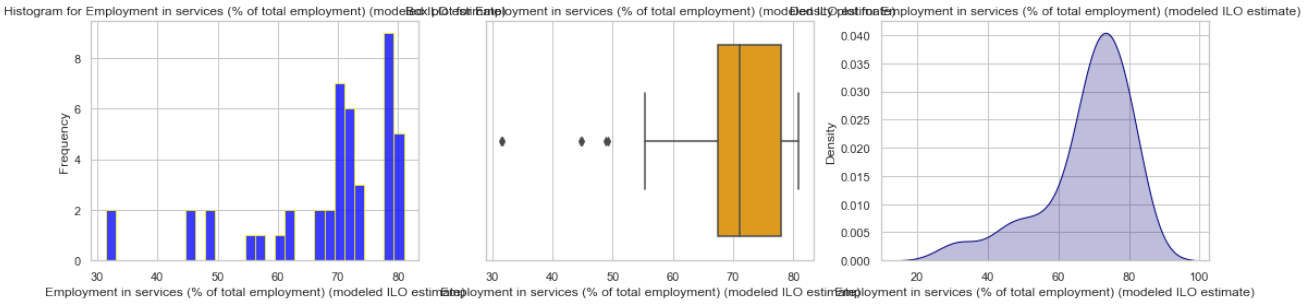

Histogram for GNI growth (annual %)
Frequency
GNI growth (annual %)
Box plot for GNI growth (annual %)
GNI growth (annual %)
Density plot for GNI growth (annual %)
Density
GNI growth (annual %)

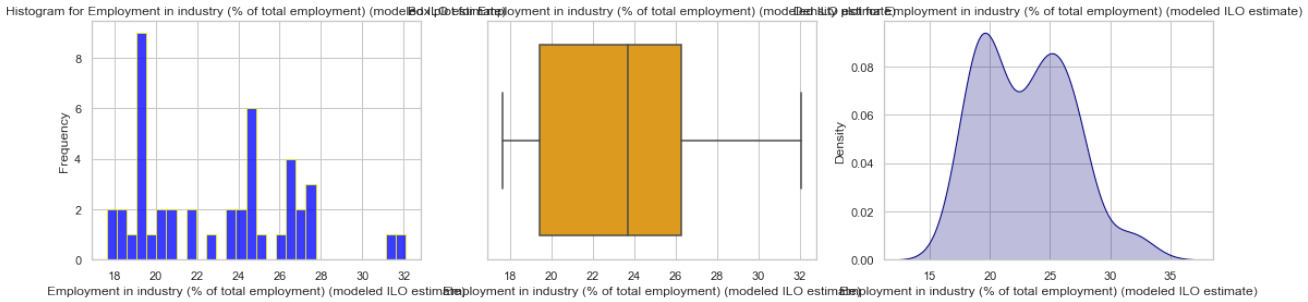

Histogram for Employment to population ratio, 15+, total (%) (national estimate)
Frequency
Employment to population ratio, 15+, total (%) (national estimate)
Boxplot for Employment to population ratio, 15+, total (%) (national estimate)
Employment to population ratio, 15+, total (%) (national estimate)
Density plot for Employment to population ratio, 15+, total (%) (national estimate)
Density
Employment to population ratio, 15+, total (%) (national estimate)

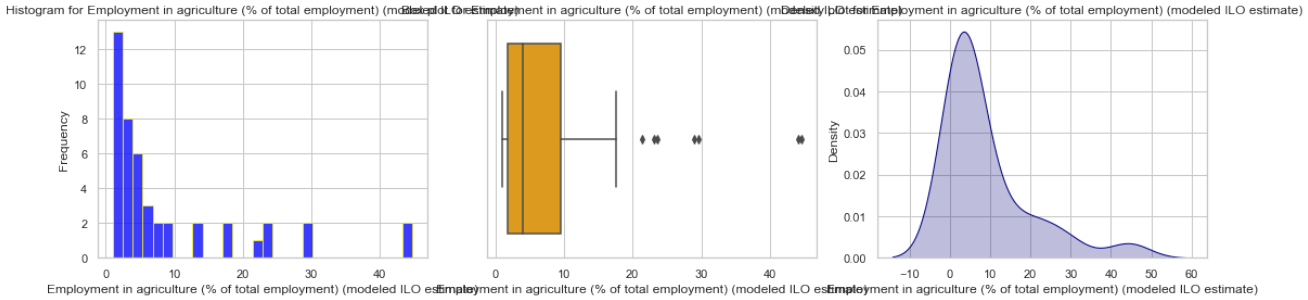

Histogram for Employment to population ratio, 15+, female (%) (national estimate)
Frequency
Employment to population ratio, 15+, female (%) (national estimate)
Boxplot for Employment to population ratio, 15+, female (%) (national estimate)
Employment to population ratio, 15+, female (%) (national estimate)
Density plot for Employment to population ratio, 15+, female (%) (national estimate)
Density
Employment to population ratio, 15+, female (%) (national estimate)


Histogram for Employment to population ratio, 15+, male (%) (national estimate)
Frequency
Employment to population ratio, 15+, male (%) (national estimate)
Boxplot for Employment to population ratio, 15+, male (%) (national estimate)
Employment to population ratio, 15+, male (%) (national estimate)
Density plot for Employment to population ratio, 15+, male (%) (national estimate)
Density
Employment to population ratio, 15+, male (%) (national estimate)

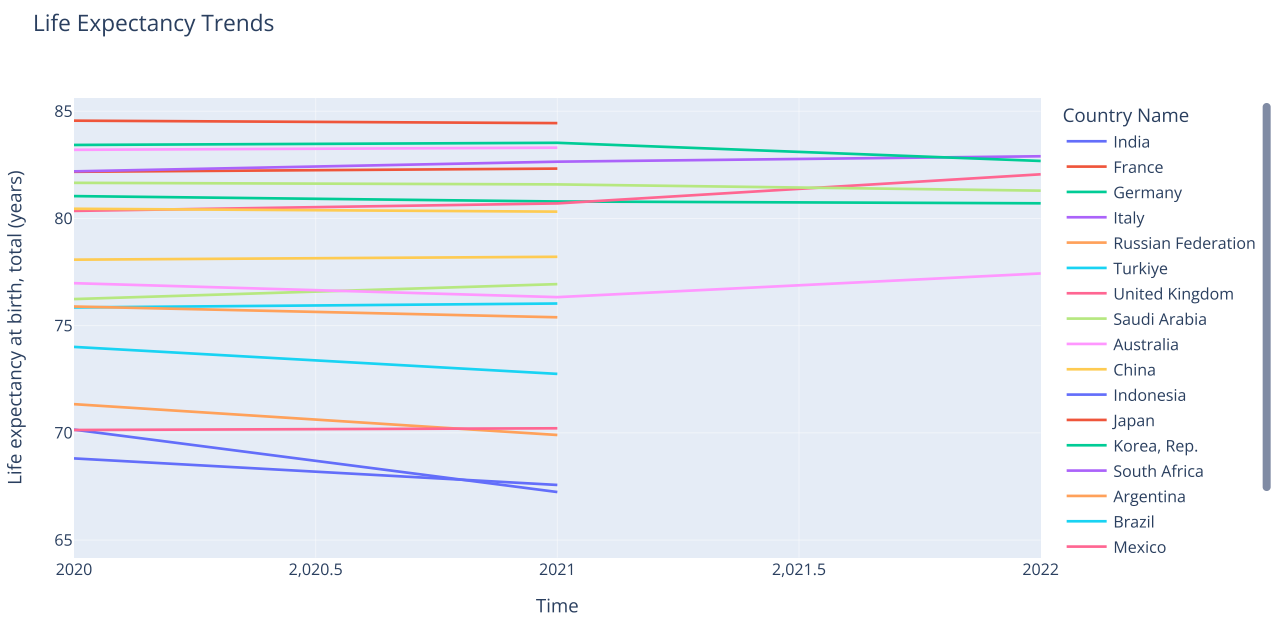

Histogram for Unemployment, total (% of total labor force) (national estimate)
Frequency
Unemployment, total (% of total labor force) (national estimate)
Boxplot for Unemployment, total (% of total labor force) (national estimate)
Unemployment, total (% of total labor force) (national estimate)
Density plot for Unemployment, total (% of total labor force) (national estimate)
Density
Unemployment, total (% of total labor force) (national estimate)


Histogram for Employment in services (% of total employment) (modeled ILO estimate)
Frequency
Employment in services (% of total employment) (modeled ILO estimate)
Boxplot for Employment in services (% of total employment) (modeled ILO estimate)
Employment in services (% of total employment) (modeled ILO estimate)
Density plot for Employment in services (% of total employment) (modeled ILO estimate)
Density
Employment in services (% of total employment) (modeled ILO estimate)

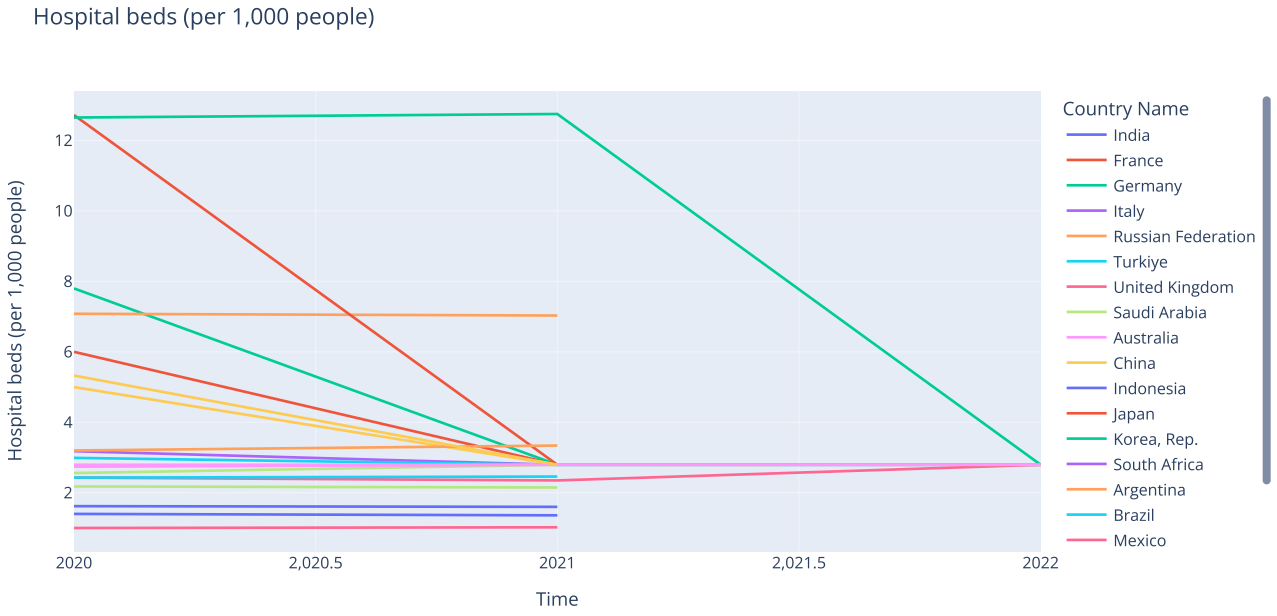




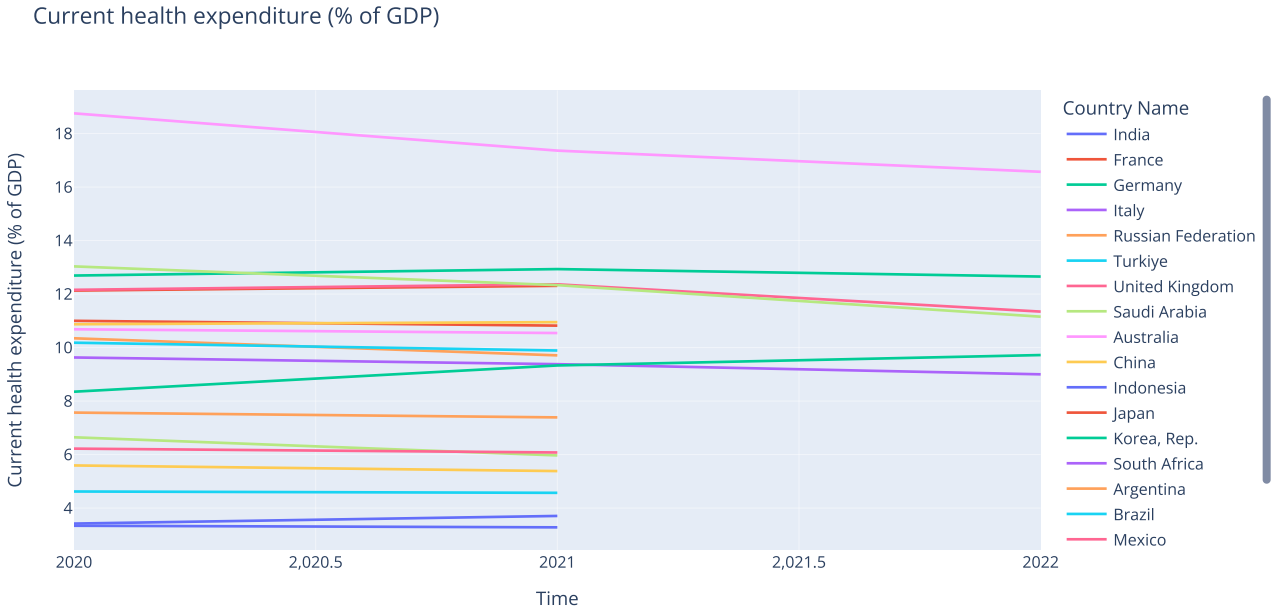
In [108]:
```
fig = px.line(df_covid_crisis, x='Time', y='Life expectancy at birth, total (years)', color='Country Name',
              title="Life Expectancy Trends")
fig.show()
```



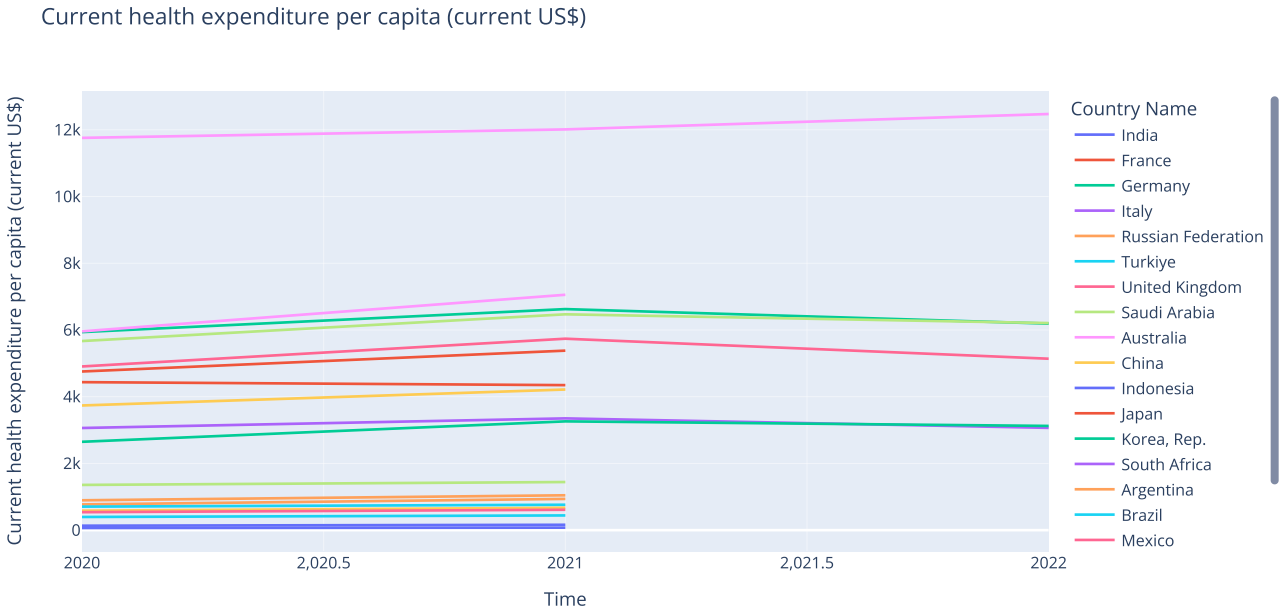
In [109]:
```
fig = px.line(df_covid_crisis, x='Time', y='Hospital beds (per 1,000 people)', color='Country Name',
              title="Hospital beds (per 1,000 people)")
fig.show()
```



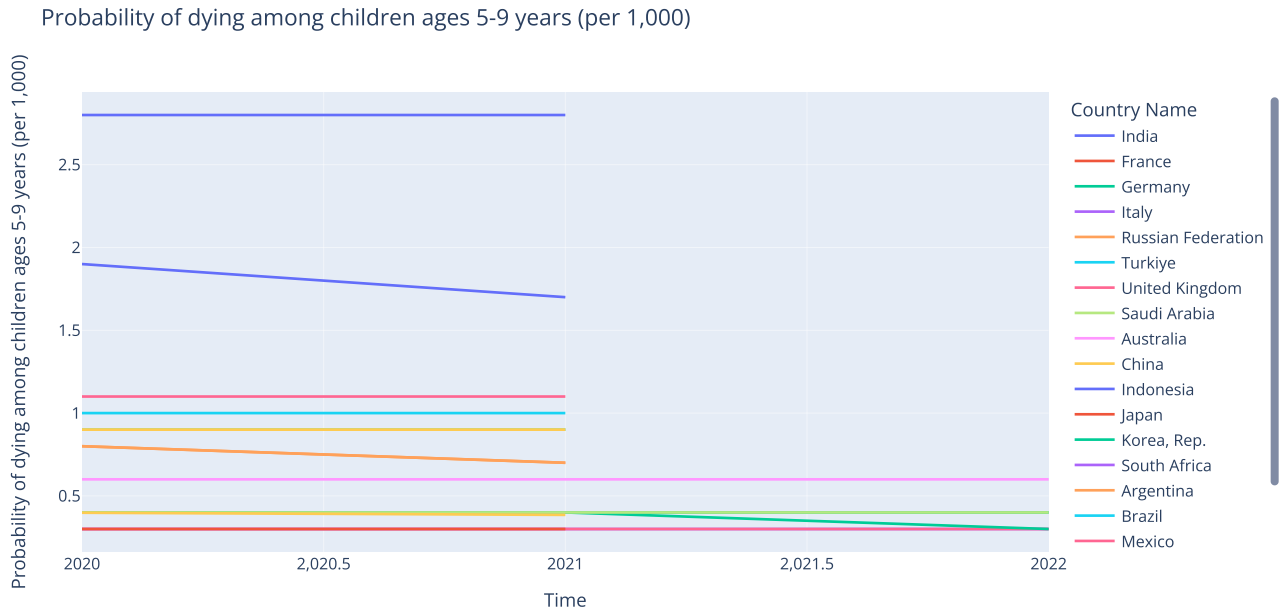
In [110]:
```
fig = px.line(df_covid_crisis, x='Time', y='Current health expenditure (% of GDP)', color='Country Name',
              title="Current health expenditure (% of GDP)")
fig.show()
```



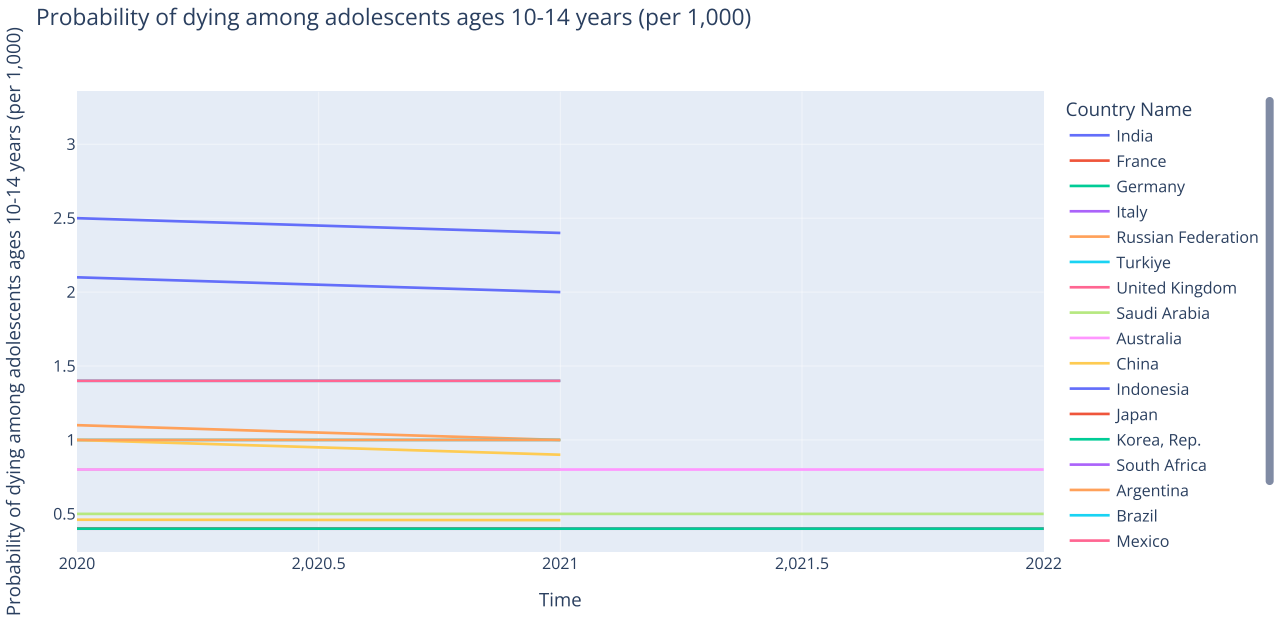
In [111]:
```
fig = px.line(df_covid_crisis, x='Time', y='Current health expenditure per capita (current US$)', color='Country Name',
              title="Current health expenditure per capita (current US$)")
fig.show()
```



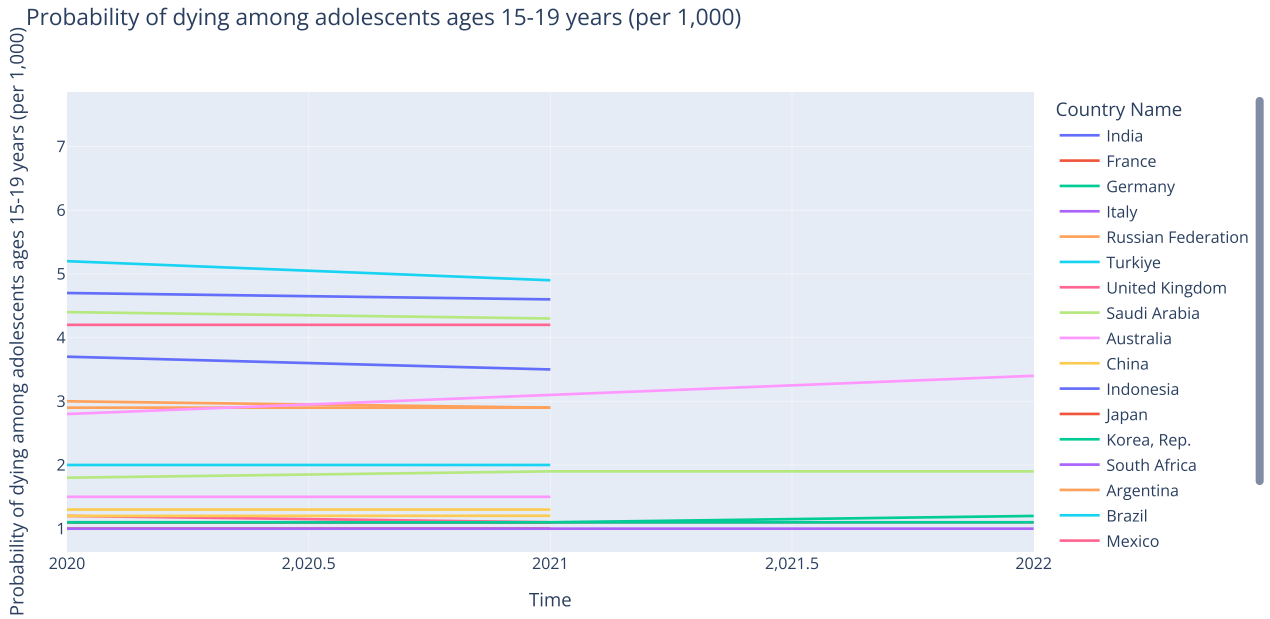
In [113]:
```
fig = px.line(df_covid_crisis, x='Time', y='Probability of dying among children ages 5-9 years (per 1,000)', color='Country Name'
              title="Probability of dying among children ages 5-9 years (per 1,000)")
fig.show()
```



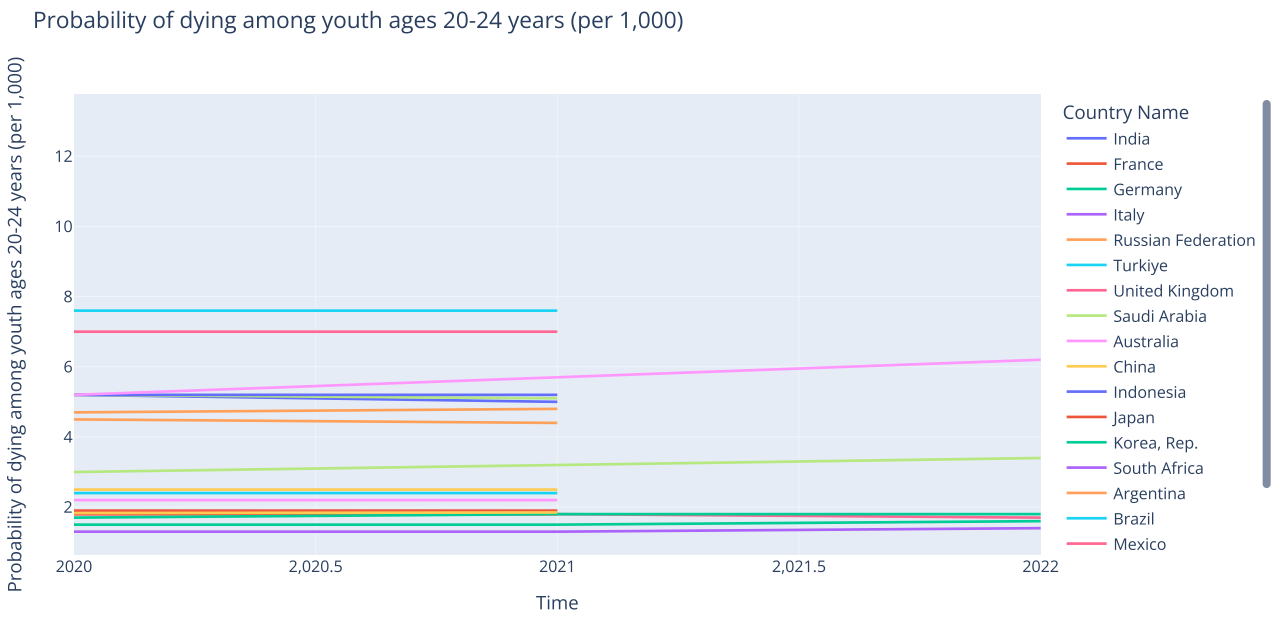
In [118]:
```
fig = px.line(df_covid_crisis, x='Time', y='Probability of dying among adolescents ages 10-14 years (per 1,000)', color='Country
              title="Probability of dying among adolescents ages 10-14 years (per 1,000)")
fig.show()
```



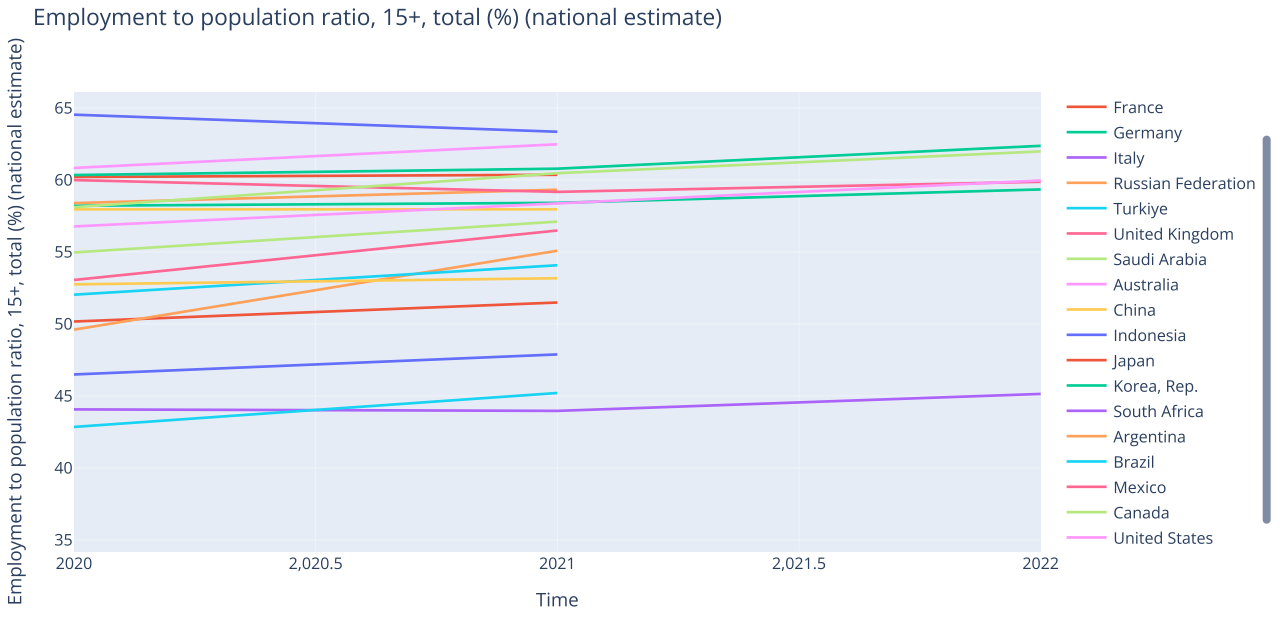
In [119]:
```
fig = px.line(df_covid_crisis, x='Time', y='Probability of dying among adolescents ages 15-19 years (per 1,000)', color='Country
              title="Probability of dying among adolescents ages 15-19 years (per 1,000)")
fig.show()
```



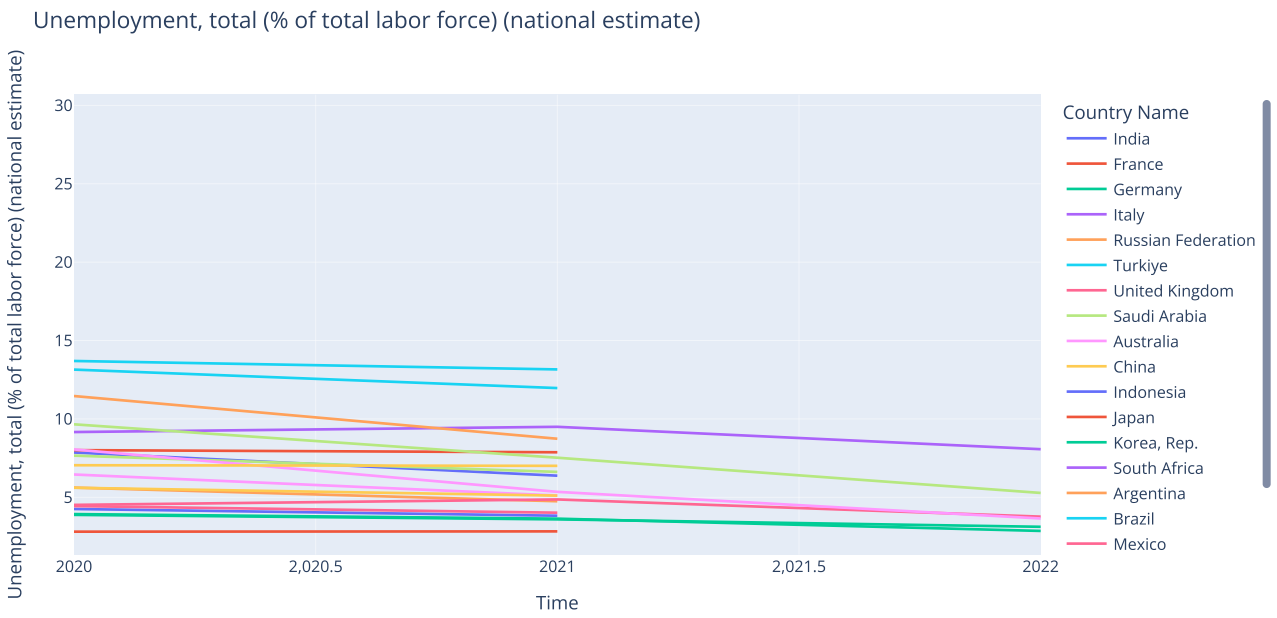
In [120]:
```
fig = px.line(df_covid_crisis, x='Time', y='Probability of dying among youth ages 20-24 years (per 1,000)', color='Country Name',
              title="Probability of dying among youth ages 20-24 years (per 1,000)")
fig.show()
```



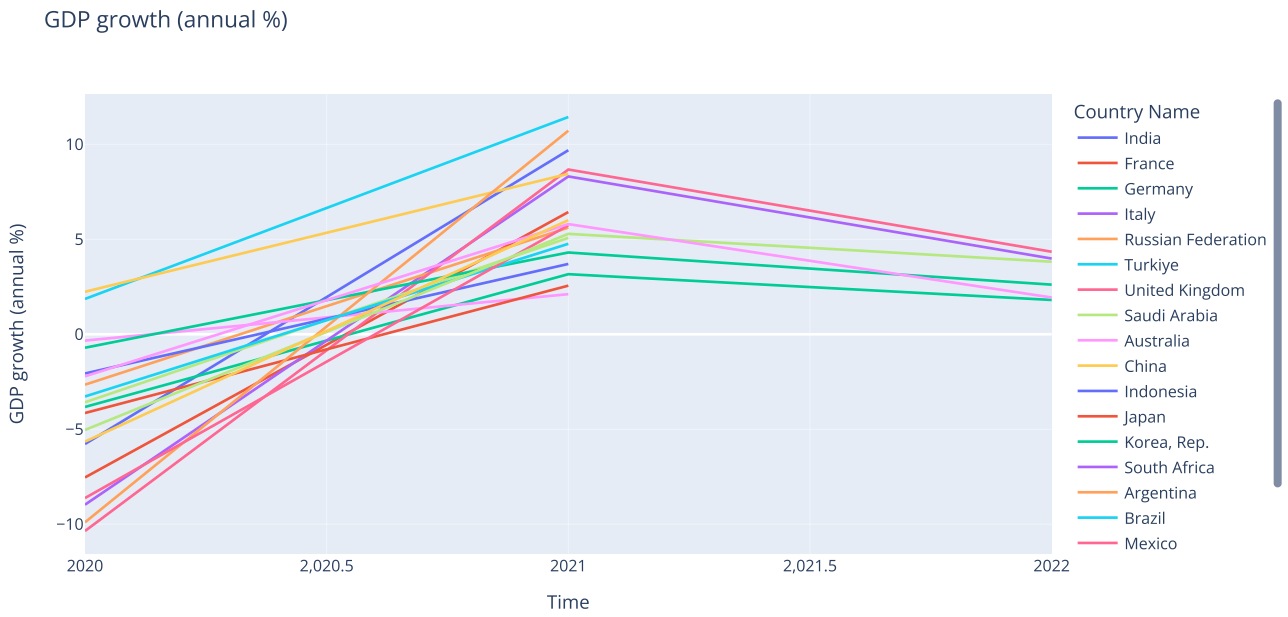
In [114]:
```
fig = px.line(df_covid_crisis, x='Time', y='Employment to population ratio, 15+, total (%) (national estimate)', color='Country N
              title="Employment to population ratio, 15+, total (%) (national estimate)")
fig.show()
```



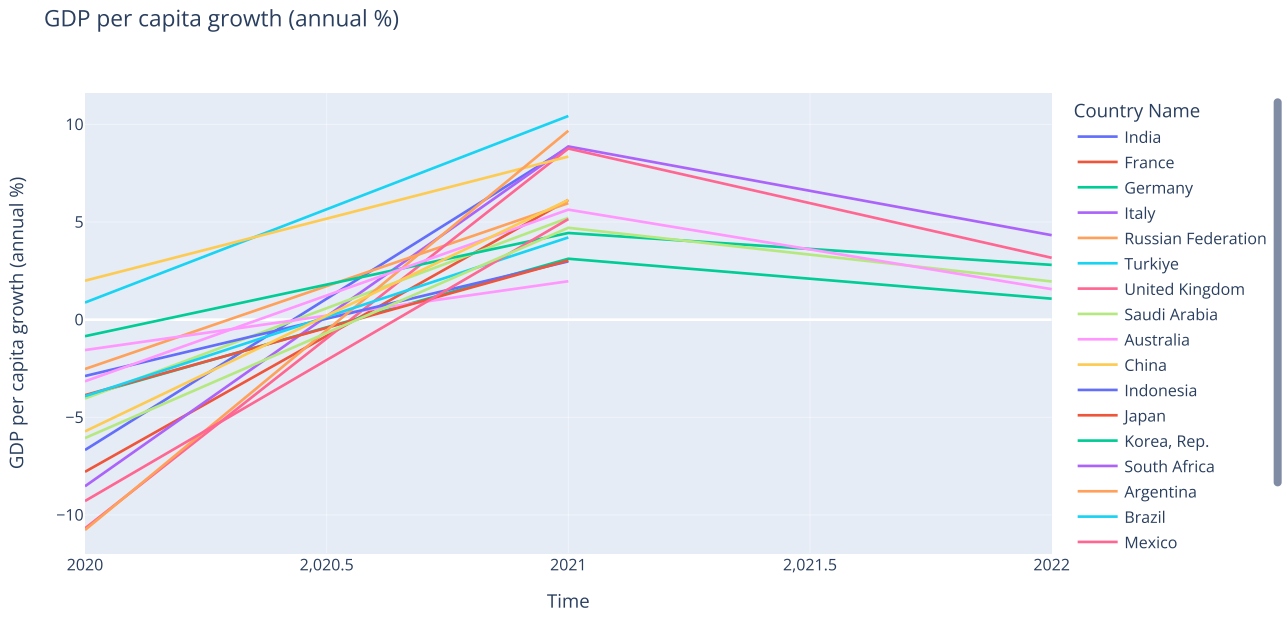
In [115]:
```
fig = px.line(df_covid_crisis, x='Time', y='Unemployment, total (% of total labor force) (national estimate)', color='Country Nam
              title="Unemployment, total (% of total labor force) (national estimate)")
fig.show()
```



In [116]:
```
fig = px.line(df_covid_crisis, x='Time', y='GDP growth (annual %)', color='Country Name',
              title="GDP growth (annual %)")
fig.show()
```



In [117]:
```
fig = px.line(df_covid_crisis, x='Time', y='GDP per capita growth (annual %)', color='Country Name',
              title="GDP per capita growth (annual %)")
fig.show()
```



In [ ]: