

Project: Investigate a Dataset (Database_Gapminder_World)

Table of Contents

- [Introduction](#)
- [Data Wrangling](#)
- [Exploratory Data Analysis](#)
- [Conclusions](#)

Introduction

Foreign debt (% GNI): Total external debt stocks to gross national income. Total external debt is debt owned to nonresidents repayable in currency, goods , or services. Total external debt is the sum of public , publicly guaranteed and private nonguaranteed long-term debt, use of IMF credit and short-term debt

Inflation (annual %) : inflation `as measured by the annual growth rateof the GDP implicit deflator shows the rateof price change in the economy as a whole. The GDP implicit deflator is the ratio of GDP in current local currency to GDP in Constant local currency

Investment (% of GDP) also known as Gross capital formation: consist of outlays on additions to the fixed assets of the economy plus net changes in level of inventories. Fixed assets include land improvements, plant, machinery, equipment and purchases, railways, construction of roads and the like, including schools, offices, hospitals, private resedential dwellings, and commercial and industrial buildings. Gross capital formation is seen as a major component of domestic investment

Trade balance (US\$ not inflation-adjusted): External balance on goods and services (formerly resource balance) equals exports of goods and service minus imports of goods and services. Data are in current US dollars

Research Questions (1)

1. Countries within organisation and regions with the highest and lowest inflation in the last 23 years.

1b. Which country within the selected Sub Saharan countries have the highest foreign debt

1c. Countries within the selected organisations and regions with the highest and lowest Trade Balance

Research Questions (2)

2a. Does inflation has any effect on Foreign debt

2b. Do countries with high foreign debt have high Gross capital formation

2c. Is there any substantial effect of foreign debt on Trade balance

2d. Does inflation have any effect or impact on Gross capital formation

2e. Countries and regions with the highest Trade Balance

2f. Effect of Gross capital formation on Trade Balance

Research Questions (3):

3a. Does the relationship between inflation and Foreign_debt vary with countries

3b. Do all countries involved in the analysis follow the same relationship that exist with Trade_balance and Gross capital formation

3c. Is the impact of inflation on Gross capital formation the same for all countries involved in the analysis

3d. Does inflation have a general effect on Foreign_debt

3e. Is the effect of Foreign_debt on Trade_balance the same for all countries

```
In [1]: ### import packages
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
from matplotlib.widgets import Slider, Button
%matplotlib inline
pd.set_option('display.max_columns', None)
```

Data Wrangling

In this section of the report, we will load in the data, check for cleanliness, and then trim and clean the dataset for analysis. Create new dataframes for selected region and organisation in this case G7 and Sub Saharan Countries .We are limiting our analysis from 1997 to 2020

General Properties

```
In [2]: # Load your data and print out a few lines. Perform operations to inspect data

## Load the f\the debt_to_foreigners debt
Foreign_debt= pd.read_csv('debt_to_foreigners.csv')

## Load the inflation data
Inflation= pd.read_csv('inflation_annual_percent.csv')

## Load the investment data
Gross_capital_formation = pd.read_csv('investments_percent_of_gdp.csv')

## Load the trade trade_balance data
Trade_Balance = pd.read_csv('trade_balance_percent_of_gdp.csv')
```

In []:

In [3]: Foreign_debt.head()

Out[3]:

	country	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	19
0	Afghanistan	NaN	N												
1	Angola	NaN	N												
2	Albania	NaN	N												
3	Argentina	19.4	18.9	20.2	14.0	10.7	14.9	18.6	20.7	23.4	30.9	35.8	46.6	55.3	4
4	Armenia	NaN	N												



In [4]: Inflation.head()

Out[4]:

	country	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	19
0	Aruba	NaN	N												
1	Afghanistan	NaN	N												
2	Angola	NaN	N												
3	Albania	NaN	N												
4	Andorra	NaN	7.85	8.52	11.8	1									



In [5]: Gross_capital_formation.head()

Out[5]:

	country	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972
0	Aruba	NaN												
1	Afghanistan	16.1	16.6	19.1	14.2	13.9	11.3	8.41	5.18	6.47	6.47	5.46	5.46	6.55
2	Angola	NaN												
3	Albania	NaN												
4	United Arab Emirates	NaN												

In [6]: Trade_Balance.head(11)

Out[6]:

	country	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972
0	Aruba	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	Afghanistan	-2.89	-3.64	-4.47	-7.69	-9.17	-10.2	-10	-7.44	-6.31	-4.89	-2.16	-5	1.1
2	Angola	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	Albania	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	United Arab Emirates	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5	Argentina	0	0	-4.69	0	0	2.08	1.66	2.5	1.08	0	0.866	-0.1	0.1
6	Armenia	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
7	American Samoa	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
8	Antigua and Barbuda	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
9	Australia	-1.07	-2.62	1.31	-0.827	1.17	-2.04	-2.19	-1.01	-2.18	-1.38	-0.244	-0.1	0.1
10	Austria	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	-0.0769	-0.1

```
In [7]: def Report(data):

    ### print the shape of the data
    print(f"There are {data.shape[0]} rows and {data.shape[1]} columns in the dataset")

    ### print the column names in a list
    print("\nFeatures :\n\n", data.columns.tolist())

    ### print the number of missing values in the dataset
    print("\nMissing values :\n\n", data.isnull().sum().sort_values(ascending=True))

    ### print the percentage of missing values of each individual features in the dataset
    print("\nPercentage of Missing Values :\n\n", round(data.isna().sum()/data.shape[0]*100, 2))

    print("\n The data type of the dataset :\n\n", data.info())

    return data.describe().T
```

```
In [8]: ## understanding the trade balance dataset
```

```
In [9]: Trade_Balance['country'].unique()
```

```
Out[9]: array(['Aruba', 'Afghanistan', 'Angola', 'Albania',
       'United Arab Emirates', 'Argentina', 'Armenia', 'American Samoa',
       'Antigua and Barbuda', 'Australia', 'Austria', 'Azerbaijan',
       'Burundi', 'Belgium', 'Benin', 'Burkina Faso', 'Bangladesh',
       'Bulgaria', 'Bahrain', 'Bahamas', 'Bosnia and Herzegovina',
       'Belarus', 'Belize', 'Bermuda', 'Bolivia', 'Brazil', 'Barbados',
       'Brunei', 'Bhutan', 'Botswana', 'Central African Republic',
       'Canada', 'Switzerland', 'Chile', 'China', "Cote d'Ivoire",
       'Cameroon', 'Congo, Dem. Rep.', 'Congo, Rep.', 'Colombia',
       'Comoros', 'Cape Verde', 'Costa Rica', 'Cuba', 'Cayman Islands',
       'Cyprus', 'Czech Republic', 'Germany', 'Djibouti', 'Dominica',
       'Denmark', 'Dominican Republic', 'Algeria', 'Ecuador', 'Egypt',
       'Eritrea', 'Spain', 'Estonia', 'Ethiopia', 'Finland', 'Fiji',
       'France', 'Faeroe Islands', 'Micronesia, Fed. Sts.', 'Gabon',
       'United Kingdom', 'Georgia', 'Ghana', 'Guinea', 'Gambia',
       'Guinea-Bissau', 'Equatorial Guinea', 'Greece', 'Greenland',
       'Guatemala', 'Guam', 'Guyana', 'Hong Kong, China', 'Honduras',
       'Croatia', 'Haiti', 'Hungary', 'Indonesia', 'India', 'Ireland',
       'Iran', 'Iraq', 'Iceland', 'Israel', 'Italy', 'Jamaica', 'Jordan',
       'Japan', 'Kazakhstan', 'Kenya', 'Kyrgyz Republic', 'Cambodia',
       'Kiribati', 'South Korea', 'Kosovo', 'Kuwait', 'Lao', 'Lebanon',
       'Libya', 'Sri Lanka', 'Lesotho', 'Lithuania', 'Luxembourg',
       'Latvia', 'Macao, China', 'Morocco', 'Moldova', 'Madagascar',
       'Maldives', 'Mexico', 'Marshall Islands', 'North Macedonia',
       'Mali', 'Malta', 'Myanmar', 'Montenegro', 'Mongolia',
       'Northern Mariana Islands', 'Mozambique', 'Mauritania',
       'Mauritius', 'Malaysia', 'Namibia', 'New Caledonia', 'Niger',
       'Nigeria', 'Nicaragua', 'Netherlands', 'Curaçao', 'Norway',
       'Nepal', 'Nauru', 'New Zealand', 'Oman', 'Pakistan', 'Panama',
       'Peru', 'Philippines', 'Papua New Guinea', 'Poland', 'Puerto Rico',
       'Portugal', 'Paraguay', 'Palestine', 'French Polynesia', 'Qatar',
       'Romania', 'Russia', 'Rwanda', 'Saudi Arabia', 'Sudan', 'Senegal',
       'Singapore', 'Solomon Islands', 'Sierra Leone', 'El Salvador',
       'San Marino', 'Somalia', 'Serbia', 'South Sudan', 'Suriname',
       'Slovak Republic', 'Slovenia', 'Sweden', 'Eswatini',
       'Sint Maarten (Dutch part)', 'Seychelles', 'Syria', 'Chad', 'Togo',
       'Thailand', 'Tajikistan', 'Turkmenistan', 'Timor-Leste', 'Tonga',
       'Tunisia', 'Turkey', 'Tanzania', 'Uganda', 'Ukraine', 'Uruguay',
       'United States', 'Uzbekistan', 'Venezuela',
       'Virgin Islands (U.S.)', 'Vietnam', 'Vanuatu', 'Samoa', 'Yemen',
       'South Africa', 'Zambia', 'Zimbabwe'], dtype=object)
```

```
In [10]: Trade_Balance['country'].value_counts()
```

```
Out[10]: Fiji          1  
Belize        1  
American Samoa  1  
Macao, China    1  
Uzbekistan     1  
..  
Slovak Republic 1  
Kiribati       1  
Iran           1  
Eswatini        1  
Tonga          1  
Name: country, Length: 197, dtype: int64
```

```
In [11]: ### Trade balance has a total of 197 countries
```

we are creating a new dataframe that contains the trade_balance of G7 countries

G7 countries are a group of the world largest, advanced and developd economies which accounts for up to 45% of the Global economy

G7 countries includes:

France

UK

US

Japan

Italy

Germany

Canada

we are creating a new dataframe that contains the trade_balance of Sub-Saharan Africa countries

Sub-Saharan Africa countries are countries that lies south of the Sahara. It consists of the countries varying from underdeveloped, developing and developed economies which accounts for up to 2.5% of the Global economy

Sub-Saharan Africa countries includes:

1. Angola

2. Benin

3. Burkina Faso

4. Botswana

5. Cape Verde

6. Cameroon

7. Central African Republic (C.A.R)

8. Chad

9. Comoros

10. Congo (DRC)

11. Congo (Rep)

12. Cote d'Ivoire

13. Equatorial Guinea

14. Guinea

15. Eritrea

16. Eswatini

17. Ethiopia

18. Gabon

19. Guinea-Bissau

20. Gambia

21. Ghana

22. Kenya

23. Lesotho

24. Liberia

25. Madagascar

26. Malawi

27. Mali

28. Mauritania

29. Mauritius

30. Mozambique

31. Namibia

32. Niger

33. Nigeria

34. Rwanda

35. Sao Tome And Principe

36. Senegal

37. Seychelles

38. Sierra Leone

39. Somalia

40. South Africa

41. South Sudan

42. Sudan

43. Tanzania

44. Togo

45. Uganda

46. Zambia

47. Zimbabwe

```
In [12]: ## create a new dataframe that consist only G7 countries
G7_trade_balance=Trade_Balance.query("country in ('France', 'Germany', 'Canada',
## store G7 countries
G7_Countries= G7_trade_balance.country

## we are only choosing columns from year 1997 to 2020
## filling nan values with mean, median or mode is quite dangerous if we were to
G7_Trade_balance= pd.concat([pd.DataFrame(G7_Countries), G7_trade_balance.iloc[:,
```

```
In [13]: ## Create a new dataframe of G7 countries where country, years and Trade_balance
G7_Trade_balance= pd.melt(frame=G7_Trade_balance, id_vars='country', value_vars=(
    var_name='years', value_name='Trade_balance')
```

In [14]: G7_Trade_balance

Out[14]:

	country	years	Trade_balance
0	Canada	1997	1.93
1	Germany	1997	1.08
2	France	1997	3.01
3	United Kingdom	1997	0.316
4	Italy	1997	3.71
...
163	France	2020	-2.02
164	United Kingdom	2020	0.133
165	Italy	2020	3.69
166	Japan	2020	-0.0168
167	United States	2020	-3.11

168 rows × 3 columns

In [15]: *## create a dataframe for Trade_balance of only Sub-Saharan countries*

```
Sub_Sahara_trade_balance=Trade_Balance.query("country in ('Cameroon', 'Congo, Dem. Rep.', 'Cote d'Ivoire', 'Eritrea', 'Ethiopia', 'Ghana', 'Kenya', 'Liberia', 'Malawi', 'Mali', 'Mauritania', 'Niger', 'Nigeria', 'Rwanda', 'Sao Tome and Principe', 'Senegal', 'South Africa', 'Togo', 'Uganda')")  
Sub_Saharan_Countries= Sub_Sahara_trade_balance.country  
  
## we are only choosing columns from year 1997 to 2020  
## filling nan values with mean, median or mode is quite dangerous if we were to  
## use it on a large dataset like this one  
Sub_Sahara_Trade_balance= pd.concat([pd.DataFrame(Sub_Saharan_Countries), Sub_Sahara_trade_balance[['years', 'Trade_balance']]])
```

In [16]: *## Create a new dataframe of Sub-Saharan countries where country, years and Trade_balance are in separate columns*

```
Sub_Saharan_Trade_balance= pd.melt(frame=Sub_Sahara_Trade_balance, id_vars='country', value_vars='years', var_name='years', value_name='Trade_balance')
```

In [17]: Sub-Saharan_Trade_balance

Out[17]:

	country	years	Trade_balance
0	Angola	1997	NaN
1	Burundi	1997	-4.6
2	Benin	1997	-4.36
3	Burkina Faso	1997	-15.8
4	Cameroon	1997	4.83
...
955	Tanzania	2020	-1.01
956	Uganda	2020	-6.17
957	South Africa	2020	4.43
958	Zambia	2020	14.3
959	Zimbabwe	2020	-1.6

960 rows × 3 columns

In [18]: Report(Trade_Balance)

There are 197 rows and 62 columns in the dataset

Features :

```
['country', '1960', '1961', '1962', '1963', '1964', '1965', '1966', '1967',
'1968', '1969', '1970', '1971', '1972', '1973', '1974', '1975', '1976', '197
7', '1978', '1979', '1980', '1981', '1982', '1983', '1984', '1985', '1986',
'1987', '1988', '1989', '1990', '1991', '1992', '1993', '1994', '1995', '199
6', '1997', '1998', '1999', '2000', '2001', '2002', '2003', '2004', '2005',
'2006', '2007', '2008', '2009', '2010', '2011', '2012', '2013', '2014', '201
5', '2016', '2017', '2018', '2019', '2020']
```

Missing values :

1960	134
1961	130
1962	129
1963	129
1964	128

In [19]: Report(G7_Trade_balance)

There are 168 rows and 3 columns in the dataset

Features :

```
['country', 'years', 'Trade_balance']
```

Missing values :

```
Trade_balance    0  
years           0  
country         0  
dtype: int64
```

Percentage of Missing Values :

```
Trade_balance    0.0  
years           0.0  
country         0.0  
dtype: float64  
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 168 entries, 0 to 167  
Data columns (total 3 columns):  
 #   Column      Non-Null Count  Dtype    
---  --          --          --  
 0   country     168 non-null    object   
 1   years       168 non-null    object   
 2   Trade_balance 168 non-null  object  
dtypes: object(3)  
memory usage: 4.1+ KB
```

The data type of the dataset :

None

Out[19]:

	count	unique	top	freq
country	168	7	Canada	24
years	168	24	2007	7
Trade_balance	168	154	-2.06	3

In []:

In [20]: Report(Inflation)

There are 212 rows and 61 columns in the dataset

Features :

```
['country', '1961', '1962', '1963', '1964', '1965', '1966', '1967', '1968',
'1969', '1970', '1971', '1972', '1973', '1974', '1975', '1976', '1977', '197
8', '1979', '1980', '1981', '1982', '1983', '1984', '1985', '1986', '1987',
'1988', '1989', '1990', '1991', '1992', '1993', '1994', '1995', '1996', '199
7', '1998', '1999', '2000', '2001', '2002', '2003', '2004', '2005', '2006',
'2007', '2008', '2009', '2010', '2011', '2012', '2013', '2014', '2015', '201
6', '2017', '2018', '2019', '2020']
```

Missing values :

1961	121
1962	117
1963	117
1964	117
1965	117

In [21]: *## create a new dataframe that consist of only G7 countries*

```
G7_inflation=Inflation.query("country in ('France', 'Germany', 'Canada', 'Japan')
G7_Countries_I=G7_inflation.country

G7_inflation=pd.concat([pd.DataFrame(G7_Countries_I), G7_inflation.iloc[:,37:]])
G7_inflation
```

Out[21]:

	country	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
32	Canada	NaN	0.888	3.07	4.59	2.04	0.843	1.26	2.44	1.36	1.07	-1.36
49	Germany	0.273	0.688	0.343	-0.489	1.3	1.38	1.32	1.12	0.405	0.397	1.77
63	France	0.878	0.95	0.204	1.55	2.01	2.07	1.86	1.62	1.94	2.16	2.56
67	United Kingdom	-0.349	1.6	1.34	1.69	1.85	2.06	2.74	2.57	3.14	2.85	2.8
92	Italy	2.57	2.4	1.54	1.79	3.03	3.27	3.15	2.67	2.01	2.12	2.48
95	Japan	0.504	-0.0267	-1.24	-1.34	-1.08	-1.39	-1.61	-1.12	-1.19	-0.862	-0.705
200	United States	1.79	1.13	1.44	2.24	2.19	1.58	1.86	2.69	3.11	3.03	2.69

In [22]: *## Create a new dataframe of G7 countries where country, years and inflation_rate*

```
G7_Inflation=pd.melt(frame=G7_inflation, id_vars='country', value_vars=G7_inflation
var_name='years', value_name='inflation_rate')
```

In [23]: G7_Inflation

Out[23]:

	country	years	inflation_rate
0	Canada	1997	NaN
1	Germany	1997	0.273
2	France	1997	0.878
3	United Kingdom	1997	-0.349
4	Italy	1997	2.57
...
163	France	2020	2.52
164	United Kingdom	2020	5.35
165	Italy	2020	1.17
166	Japan	2020	0.842
167	United States	2020	1.21

168 rows × 3 columns

In [24]: Sub_Sahara_inflation=Inflation.query("country in ('Cameroon', 'Congo, Dem. Rep.'
Sub_Saharan_Countries= Sub_Sahara_inflation.country

we are only choosing columns from year 1997 to 2020
filling nan values with mean, median or mode is quite dangerous if we were to
Sub_Saharan_inflation= pd.concat([pd.DataFrame(Sub_Saharan_Countries), Sub_Sahara_inflation[['years', 'inflation_rate']]])

In [25]: ## Create a new dataframe of Sub-Saharan countries where country, years and inflation_rate are columns

Sub_Saharan_Inflation= pd.melt(frame=Sub_Saharan_inflation, id_vars='country', value_vars=['years', 'inflation_rate'])

In [26]: Sub-Saharan_Inflation

Out[26]:

	country	years	inflation_rate
0	Angola	1997	95.5
1	Burundi	1997	32.4
2	Benin	1997	3.67
3	Burkina Faso	1997	1.56
4	Cameroon	1997	6.16
...
979	Tanzania	2020	0.344
980	Uganda	2020	2.72
981	South Africa	2020	5.27
982	Zambia	2020	13.7
983	Zimbabwe	2020	559

984 rows × 3 columns

In [27]: Report(Foreign_debt)

There are 123 rows and 52 columns in the dataset

Features :

```
['country', '1970', '1971', '1972', '1973', '1974', '1975', '1976', '1977',
'1978', '1979', '1980', '1981', '1982', '1983', '1984', '1985', '1986', '198
7', '1988', '1989', '1990', '1991', '1992', '1993', '1994', '1995', '1996',
'1997', '1998', '1999', '2000', '2001', '2002', '2003', '2004', '2005', '200
6', '2007', '2008', '2009', '2010', '2011', '2012', '2013', '2014', '2015',
'2016', '2017', '2018', '2019', '2020']
```

Missing values :

1970	58
1971	58
1972	58
1975	58
1973	57
1974	57
1976	55

In [28]: `Foreign_debt['country'].unique()`

Out[28]: `array(['Afghanistan', 'Angola', 'Albania', 'Argentina', 'Armenia', 'Azerbaijan', 'Burundi', 'Benin', 'Burkina Faso', 'Bangladesh', 'Bulgaria', 'Bosnia and Herzegovina', 'Belarus', 'Belize', 'Bolivia', 'Brazil', 'Bhutan', 'Botswana', 'Central African Republic', 'China', "Cote d'Ivoire", 'Cameroon', 'Congo, Dem. Rep.', 'Congo, Rep.', 'Colombia', 'Comoros', 'Cape Verde', 'Costa Rica', 'Djibouti', 'Dominica', 'Dominican Republic', 'Algeria', 'Ecuador', 'Egypt', 'Eritrea', 'Ethiopia', 'Fiji', 'Gabon', 'Georgia', 'Ghana', 'Guinea', 'Gambia', 'Guinea-Bissau', 'Grenada', 'Guatemala', 'Guyana', 'Honduras', 'Haiti', 'Indonesia', 'India', 'Iran', 'Jamaica', 'Jordan', 'Kazakhstan', 'Kenya', 'Kyrgyz Republic', 'Cambodia', 'Kosovo', 'Lao', 'Lebanon', 'Liberia', 'St. Lucia', 'Sri Lanka', 'Lesotho', 'Morocco', 'Moldova', 'Madagascar', 'Maldives', 'Mexico', 'North Macedonia', 'Mali', 'Myanmar', 'Montenegro', 'Mongolia', 'Mozambique', 'Mauritania', 'Mauritius', 'Malawi', 'Niger', 'Nigeria', 'Nicaragua', 'Nepal', 'Pakistan', 'Panama', 'Peru', 'Philippines', 'Papua New Guinea', 'Paraguay', 'Romania', 'Russia', 'Rwanda', 'Sudan', 'Senegal', 'Solomon Islands', 'Sierra Leone', 'El Salvador', 'Somalia', 'Serbia', 'Sao Tome and Principe', 'Eswatini', 'Syria', 'Chad', 'Togo', 'Thailand', 'Tajikistan', 'Turkmenistan', 'Timor-Leste', 'Tonga', 'Tunisia', 'Turkey', 'Tanzania', 'Uganda', 'Ukraine', 'Uzbekistan', 'St. Vincent and the Grenadines', 'Venezuela', 'Vietnam', 'Vanuatu', 'Samoa', 'Yemen', 'South Africa', 'Zambia', 'Zimbabwe'], dtype=object)`

In [29]: `## create a new dataframe of G7 countries for foreign debt`

```
G7_foreign_debt=Foreign_debt.query("country in ('France', 'Germany', 'Canada', 'United Kingdom', 'Italy', 'Japan', 'United States')")  
G7_foreign_debt
```

Out[29]:

country	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983

As we can see no G7 countries are on the foreign debt list

In [30]: `## create a dataframe for sub saharan countries with foreign debt`

```
Sub_Sahara_foreign_debt=Foreign_debt.query("country in ('Cameroon', 'Congo, Dem. Rep.', 'Congo, Rep.', 'Ivory Coast', 'Kenya', 'Liberia', 'Malawi', 'Mali', 'Mauritania', 'Niger', 'Nigeria', 'Rwanda', 'Senegal', 'Tanzania', 'Uganda', 'Zambia', 'Zimbabwe')")  
Sub-Saharan_Countries= Sub_Sahara_foreign_debt.country  
  
## we are only choosing columns from year 1997 to 2020  
## filling nan values with mean, median or mode is quite dangerous if we were to  
## use it for prediction  
Sub_Saharan_foreign_debt= pd.concat([pd.DataFrame(Sub_Saharan_Countries), Sub_Sahara_foreign_debt[1997:2020].mean(), axis=1])
```

In [31]: Sub-Saharan_Foreign_debt= pd.melt(frame=Sub-Saharan_foreign_debt, id_vars='country', value_vars=Sub-Saharan_foreign_debt.columns.to_list()[1:])

In [32]: Sub-Saharan_Foreign_debt

Out[32]:

	country	years	Foreign_debt(\$)
0	Angola	1997	148.0
1	Burundi	1997	111.0
2	Benin	1997	63.0
3	Burkina Faso	1997	53.5
4	Cameroon	1997	105.0
...
883	Tanzania	2020	41.3
884	Uganda	2020	46.5
885	South Africa	2020	51.8
886	Zambia	2020	171.0
887	Zimbabwe	2020	73.0

888 rows × 3 columns

In [33]: Report(Gross_capital_formation)

There are 189 rows and 62 columns in the dataset

Features :

```
['country', '1960', '1961', '1962', '1963', '1964', '1965', '1966', '1967', '1968', '1969', '1970', '1971', '1972', '1973', '1974', '1975', '1976', '1977', '1978', '1979', '1980', '1981', '1982', '1983', '1984', '1985', '1986', '1987', '1988', '1989', '1990', '1991', '1992', '1993', '1994', '1995', '1996', '1997', '1998', '1999', '2000', '2001', '2002', '2003', '2004', '2005', '2006', '2007', '2008', '2009', '2010', '2011', '2012', '2013', '2014', '2015', '2016', '2017', '2018', '2019', '2020']
```

Missing values :

1960	127
1961	126
1962	125
1963	124
1964	123

```
In [34]: G7_GNI=Gross_capital_formation.query("country in ('France', 'Germany', 'Canada')", G7_GNI
G7_Countries_GNI= G7_GNI.country

G7_GNI= pd.concat([pd.DataFrame(G7_Countries_GNI), G7_GNI.iloc[:,38:]], axis=1)
G7_GNI
```

Out[34]:

	country	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
30	Canada	21.2	21.0	20.7	20.6	19.7	19.7	20.5	21.4	22.6	23.6	23.9	24.1	22.0	21.8
45	Germany	23.4	24.0	24.0	24.5	23.0	20.8	20.4	19.8	19.5	20.6	21.4	21.4	18.6	21.0
59	France	19.5	20.7	21.4	22.5	22.2	21.3	21.2	21.9	22.5	23.2	24.2	24.1	21.3	21.2
63	United Kingdom	17.6	18.0	17.9	18.2	18.0	18.1	17.7	17.5	17.7	18.0	18.4	17.4	14.9	11.0
86	Italy	19.5	19.7	20.2	20.9	20.8	21.6	21.2	21.3	21.2	22.0	22.2	21.8	19.5	21.0
89	Japan	31.1	29.6	28.2	28.4	27.8	25.9	25.7	25.6	26.1	26.1	25.8	25.9	22.6	21.0
181	United States	22.4	23.0	23.4	23.7	22.2	21.7	21.7	22.7	23.4	23.5	22.6	21.1	17.8	11.0

```
In [35]: Sub_Sahara_GNI=Gross_capital_formation.query("country in ('Cameroon', 'Congo', 'Dem. Rep. Congo', 'Ecuador', 'Ghana', 'Guinea', 'Kenya', 'Liberia', 'Madagascar', 'Malawi', 'Mali', 'Niger', 'Nigeria', 'Rwanda', 'Senegal', 'Togo', 'Uganda')", Sub_Saharan_Countries= Sub_Sahara_GNI.country

## we are only choosing columns from year 1997 to 2020
## filling nan values with mean, median or mode is quite dangerous if we were to do it for all columns

Sub_Sahara_GNI= pd.concat([pd.DataFrame(Sub_Saharan_Countries), Sub_Sahara_GNI.iloc[:,38:]], axis=1)
```

In []:

```
In [36]: ## Create a dataframe of G7 countries where country, years and Investment_perceptor are columns

Investment_G7= pd.melt(frame=G7_GNI, id_vars='country',
value_vars=G7_GNI.columns.tolist()[1:], var_name='Year')
```

In [37]: Investment_G7

Out[37]:

	country	years	Investment_percent_GDP
0	Canada	1997	21.2
1	Germany	1997	23.4
2	France	1997	19.5
3	United Kingdom	1997	17.6
4	Italy	1997	19.5
...
163	France	2020	23.8
164	United Kingdom	2020	16.7
165	Italy	2020	17.6
166	Japan	2020	25.5
167	United States	2020	21.1

168 rows × 3 columns

In [38]: *## Create a dataframe of Sub-Saharan countries where country, years and Investment_percent_GDP are columns.*

```
Investment_SSafrica= pd.melt(frame=Sub_Sahara_GNI, id_vars='country',
                               value_vars=Sub_Sahara_GNI.columns.tolist()[1:]).dropna()
```

In [39]: Investment_SSafrica.isna().sum()

Out[39]:

country	0
years	0
Investment_percent_GDP	99
dtype: int64	

Data Cleaning (Replace this with more specific notes!)

In [40]: *# After discussing the structure of the data and any problems that need to be cleaned, perform those cleaning steps in the second part of this section.*

In [41]: Report(Investment_SSafrica)

There are 960 rows and 3 columns in the dataset

Features :

```
['country', 'years', 'Investment_percent_GDP']
```

Missing values :

```
Investment_percent_GDP      99
years                      0
country                     0
dtype: int64
```

Percentage of Missing Values :

```
Investment_percent_GDP      10.312
years                      0.000
country                     0.000
dtype: float64
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 960 entries, 0 to 959
Data columns (total 3 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   country          960 non-null    object  
 1   years            960 non-null    object  
 2   Investment_percent_GDP  861 non-null  object  
dtypes: object(3)
memory usage: 22.6+ KB
```

The data type of the dataset :

None

Out[41]:

		count	unique	top	freq
	country	960	40	Comoros	24
	years	960	24	2019	40
	Investment_percent_GDP	861	397	18	11

we have 99 missing values from the investment_percent_GDP, which is about (10% of the entire dataset)

Investment_percent_GDP is an object type which is a major problem for our analysis that we would have to deal with

In [42]: Report(Investment_G7)

There are 168 rows and 3 columns in the dataset

Features :

```
['country', 'years', 'Investment_percent_GDP']
```

Missing values :

```
Investment_percent_GDP      0
years                      0
country                     0
dtype: int64
```

Percentage of Missing Values :

```
Investment_percent_GDP      0.0
years                      0.0
country                     0.0
dtype: float64
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 168 entries, 0 to 167
Data columns (total 3 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   country          168 non-null    object  
 1   years            168 non-null    object  
 2   Investment_percent_GDP  168 non-null  object  
dtypes: object(3)
memory usage: 4.1+ KB
```

The data type of the dataset :

None

Out[42]:

		count	unique	top	freq
	country	168	7	Canada	24
	years	168	24	2007	7
	Investment_percent_GDP	168	90	22.6	6

Investment_G7 has no missing value

but investment_percent_GDP is of object type , which we would have to correct

In [43]: Report(Sub-Saharan_Foreign_debt)

There are 888 rows and 3 columns in the dataset

Features :

```
['country', 'years', 'Foreign_debt($)']
```

Missing values :

```
Foreign_debt($)      24
years                 0
country               0
dtype: int64
```

Percentage of Missing Values :

```
Foreign_debt($)      2.703
years                 0.000
country               0.000
dtype: float64
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 888 entries, 0 to 887
Data columns (total 3 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   country          888 non-null    object  
 1   years             888 non-null    object  
 2   Foreign_debt($)  864 non-null    float64 
dtypes: float64(1), object(2)
memory usage: 20.9+ KB
```

The data type of the dataset :

None

Out[43]:

	count	mean	std	min	25%	50%	75%	max
Foreign_debt(\$)	864.0	59.057998	51.775591	4.95	25.2	44.3	73.225	506.0

Sub-Saharan_Foreign_debt has 24 missing values

foreign_debt is of object typw , which we would have to convert to float type

In [44]: Report(Sub-Saharan_Inflation)

There are 984 rows and 3 columns in the dataset

Features :

```
['country', 'years', 'inflation_rate']
```

Missing values :

```
inflation_rate    43
years            0
country          0
dtype: int64
```

Percentage of Missing Values :

```
inflation_rate    4.37
years            0.00
country          0.00
dtype: float64
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 984 entries, 0 to 983
Data columns (total 3 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   country          984 non-null    object 
 1   years            984 non-null    object 
 2   inflation_rate  941 non-null    object 
dtypes: object(3)
memory usage: 23.2+ KB
```

The data type of the dataset :

None

Out[44]:

	count	unique	top	freq
country	984	41	Comoros	24
years	984	24	2019	41
inflation_rate	941	716	14	8

Sub-Saharan_Inflation has about 4.37% missing values from its dataset

inflation_rate is of object type instead of float type

In [45]: Report(G7_Inflation)

There are 168 rows and 3 columns in the dataset

Features :

```
['country', 'years', 'inflation_rate']
```

Missing values :

```
inflation_rate    1
years            0
country          0
dtype: int64
```

Percentage of Missing Values :

```
inflation_rate    0.595
years            0.000
country          0.000
dtype: float64
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 168 entries, 0 to 167
Data columns (total 3 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   country     168 non-null    object 
 1   years       168 non-null    object 
 2   inflation_rate  167 non-null  object 
dtypes: object(3)
memory usage: 4.1+ KB
```

The data type of the dataset :

None

Out[45]:

	count	unique	top	freq
country	168	7	Canada	24
years	168	24	2007	7
inflation_rate	167	135	1.07	4

In []:

In [46]: *## a function to transform object type to float*

```
def make_float(variable):
    num= variable.translate({0xa0:None, 0x2212:'-'})
    return float(num)
```

```
In [47]: ## transform column inflation_rate to float, but drop all NaN values if present
```

```
G7_Inflation['inflation_rate']= G7_Inflation['inflation_rate'].dropna().apply(ma
```

a new dataframe consisting of the 7 biggest economy in Sub-Saharan_africa

This countries include:

Nigeria

Ghana

South Africa

Rwanda

Angola

Kenya

```
In [48]: Investment_SS7=Investment_SSafrica.query("country in ('Nigeria', 'South Africa',
```

```
<----->
```

```
In [49]: SS7_Inflation=Sub-Saharan_Inflation.query("country in ('Nigeria', 'South Africa',
```

```
<----->
```

```
In [50]: Foreign_debt_SS7=Sub-Saharan_Foreign_debt.query("country in ('Nigeria', 'South Afri
```

```
In [51]: Trade_balance_SS7=Sub-Saharan_Trade_balance.query("country in ('Nigeria', 'South Afri
```

```
In [52]: Foreign_debt_SS7.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 168 entries, 0 to 885
Data columns (total 3 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   country          168 non-null    object  
 1   years            168 non-null    object  
 2   Foreign_debt($)  168 non-null    float64 
dtypes: float64(1), object(2)
memory usage: 5.2+ KB
```

```
In [53]: Foreign_debt_SS7
```

```
Out[53]:
```

	country	years	Foreign_debt(\$)
0	Angola	1997	148.0
11	Ghana	1997	93.9
15	Kenya	1997	49.9
23	Nigeria	1997	54.5
24	Rwanda	1997	60.6
...
866	Kenya	2020	38.5
874	Nigeria	2020	16.9
875	Rwanda	2020	81.1
877	Senegal	2020	71.7
885	South Africa	2020	51.8

168 rows × 3 columns

In [54]: SS7_Inflation

Out[54]:

	country	years	inflation_rate
0	Angola	1997	95.5
11	Ghana	1997	19.5
16	Kenya	1997	11.4
25	Nigeria	1997	5.06
26	Rwanda	1997	15.6
...
959	Kenya	2020	5.18
968	Nigeria	2020	7.85
969	Rwanda	2020	8.27
971	Senegal	2020	2.34
981	South Africa	2020	5.27

168 rows × 3 columns

In [55]: *## we queried the dataframe if -30μ does exist*
 Trade_balance_SS7.query("Trade_balance == '-30μ'")

Out[55]:

	country	years	Trade_balance
744	Nigeria	2015	-30μ

In [56]: *## we replaced -30μ with the original Trade_balance for Nigeria in 2015 with 0*
 Trade_balance_SS7.Trade_balance[Trade_balance_SS7.Trade_balance=='-30μ'] = '0'

C:\Users\MAINGATE\AppData\Roaming\Python\Python37\site-packages\pandas\core\series.py:1020: SettingWithCopyWarning:
 A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
 self._where(~key, value, inplace=True)

we carried out this replacement on Trade_balance_Ss7 to make it easier to convert to float

In []:

In [57]: `## we can now drop missing values
SS7_Trade_balance=Trade_balance_SS7.dropna()`

In [58]: `## drop missing values in investment_ss7
Investment_ss7=Investment_SS7.dropna()`

In [59]: `Investment_ss7.info()`

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 165 entries, 11 to 957
Data columns (total 3 columns):
 #   Column           Non-Null Count  Dtype  
---  --  
 0   country          165 non-null    object  
 1   years             165 non-null    object  
 2   Investment_percent_GDP  165 non-null    object  
dtypes: object(3)
memory usage: 5.2+ KB
```

In [60]: `## change the object type in Investment_percent_GDP to float`

```
Investment_ss7['Investment_percent_GDP']= Investment_ss7['Investment_percent_GDP']
```

C:\Users\MAINGATE\AppData\Roaming\Python\Python37\site-packages\ipykernel_launcher.py:3: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.

Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

This is separate from the ipykernel package so we can avoid doing imports until

In [61]: `## change the object type of inflation_rate to float type using the function make_float
we can't use astype because negative signs now represented by emdash(--)`

```
SS7_Inflation['inflation_rate']= SS7_Inflation['inflation_rate'].apply(make_float)
```

C:\Users\MAINGATE\AppData\Roaming\Python\Python37\site-packages\ipykernel_launcher.py:4: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.

Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

after removing the cwd from sys.path.

In [62]: `G7_Trade_balance['Trade_balance']= G7_Trade_balance['Trade_balance'].apply(make_float)`

In [63]: G7_Trade_balance.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 168 entries, 0 to 167
Data columns (total 3 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   country          168 non-null    object  
 1   years            168 non-null    object  
 2   Trade_balance    168 non-null    float64 
dtypes: float64(1), object(2)
memory usage: 4.1+ KB
```

In [64]: SS7_Trade_balance['Trade_balance']= SS7_Trade_balance['Trade_balance'].apply(make

```
C:\Users\MAINGATE\AppData\Roaming\Python\Python37\site-packages\ipykernel_launcher.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

"""Entry point for launching an IPython kernel.

In [65]: SS7_Trade_balance.info()

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 165 entries, 11 to 957
Data columns (total 3 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   country          165 non-null    object  
 1   years            165 non-null    object  
 2   Trade_balance    165 non-null    float64 
dtypes: float64(1), object(2)
memory usage: 5.2+ KB
```

In [66]: Investment_G7['Investment_percent_GDP']= Investment_G7.Investment_percent_GDP.ast

In [67]: Investment_G7.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 168 entries, 0 to 167
Data columns (total 3 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   country          168 non-null    object  
 1   years            168 non-null    object  
 2   Investment_percent_GDP  168 non-null    float64 
dtypes: float64(1), object(2)
memory usage: 4.1+ KB
```

In []:

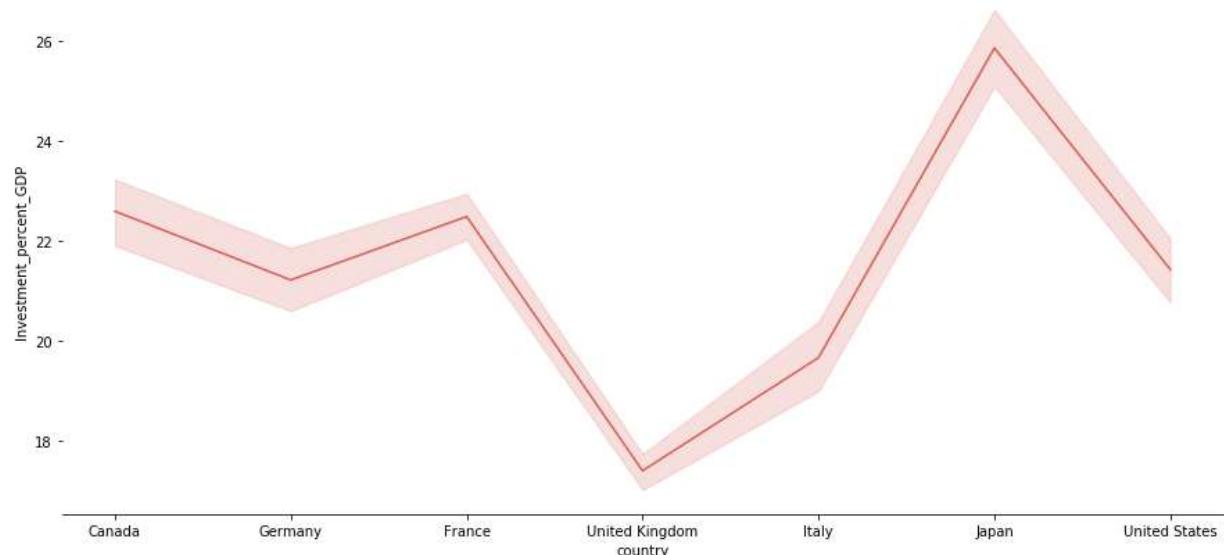
Exploratory Data Analysis

Tip: Now that you've trimmed and cleaned your data, you're ready to move on to exploration. Compute statistics and create visualizations with the goal of addressing the research questions that you posed in the Introduction section. It is recommended that you be systematic with your approach. Look at one variable at a time, and then follow it up by looking at relationships between variables.

Research Question 1 (Replace this header name!)

In [68]:

```
sns.set_palette(sns.color_palette("hls", 8))
fig, axes = plt.subplots(1,1, sharex=True, sharey=False)
fig.set_figheight(7)
fig.set_figwidth(15)
fig.set_facecolor('white')
sns.lineplot(y='Investment_percent_GDP', x= 'country', data=Investment_G7)
sns.despine(top=True, right=True, left=True)
```

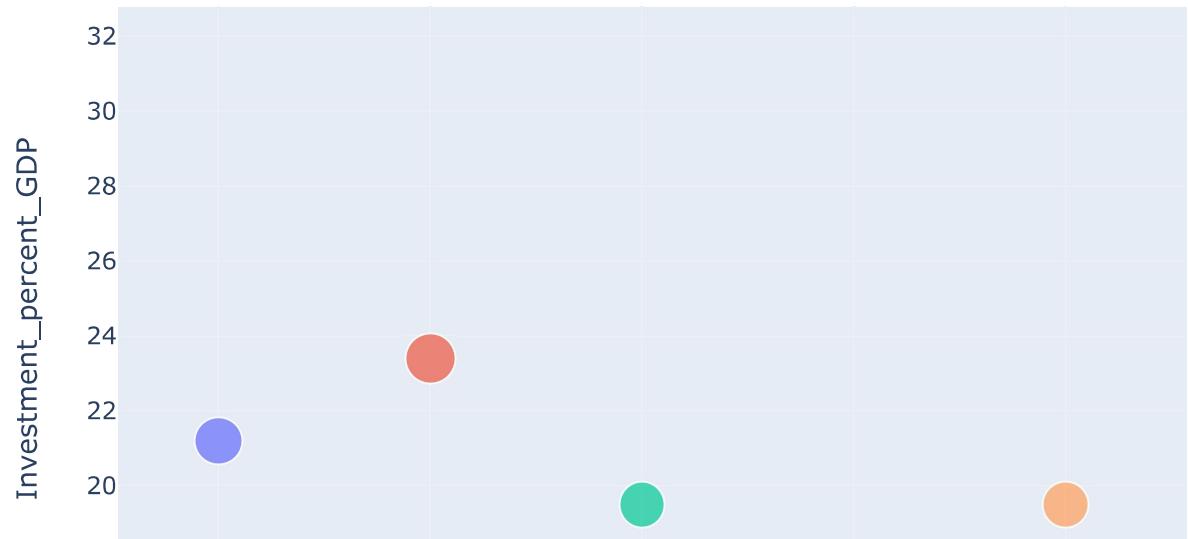


In [69]:

```
def line_plot(data_frame, variable1:str, variable2:str):

    sns.set_palette(sns.color_palette("hls", 8))
    fig, axes = plt.subplots(1,1, sharex=True, sharey=False)
    fig.set_figheight(7)
    fig.set_figwidth(15)
    fig.set_facecolor('white')
    sns.lineplot(y=variable2, x= variable1, data=data_frame)
```

```
In [70]: px.scatter(Investment_G7, x='country', y="Investment_percent_GDP", animation_fra  
size= 'Investment_percent_GDP')
```

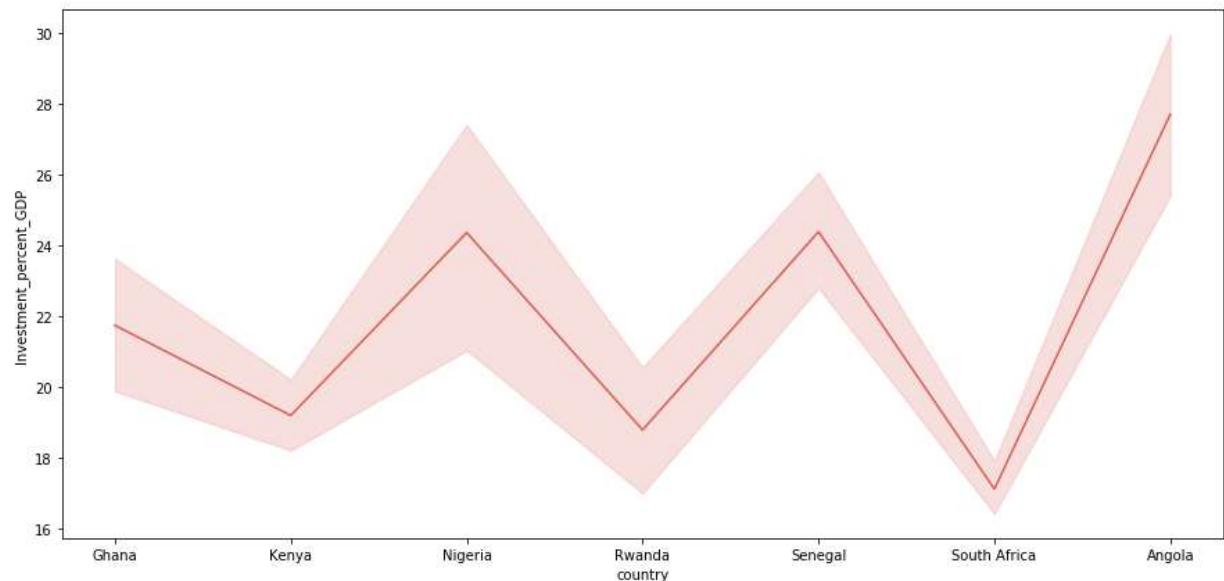


Amongst the G7 countries Japan and France had the highest Gross capital formation(investment_percent_GDP) with an average of 25% and 22% respectively

While the United Kingdom and italy have the lowest Investment_percent_GDP with an average of 17% and 30% respectively

```
In [ ]:
```

```
In [71]: line_plot(Investment_ss7, 'country', 'Investment_percent_GDP' )
```



```
In [72]: px.scatter(Investment_ss7, x='country', y="Investment_percent_GDP", animation_fra  
size= 'Investment_percent_GDP')
```

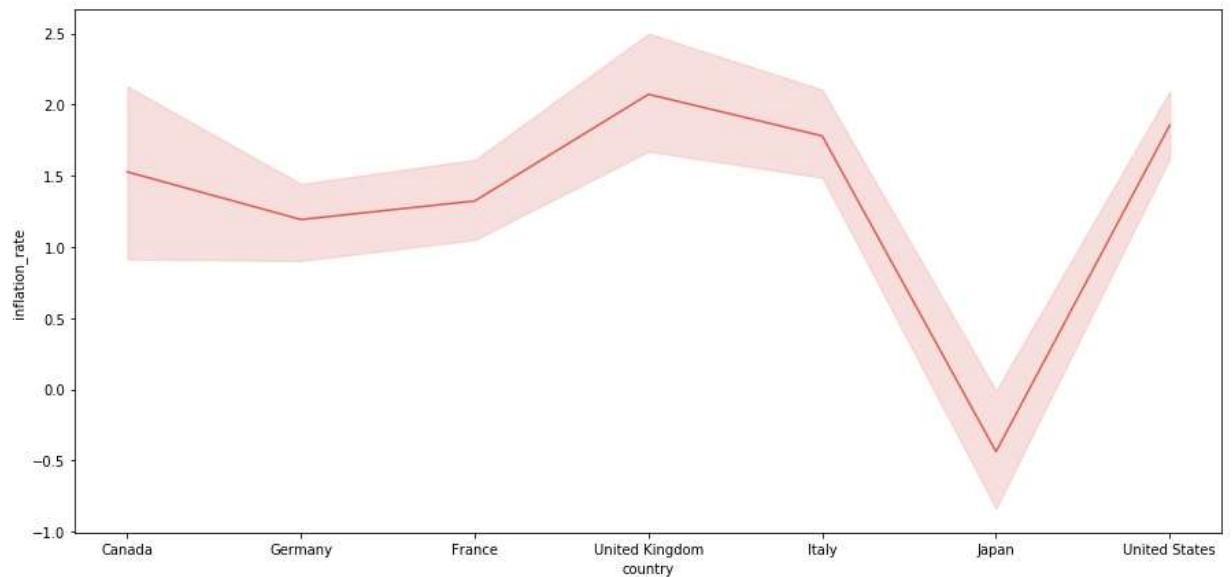


Angola, Nigeria and Senegal amongst the selected sub saharan countries have the highest Investment_percent_GDP average,with about 27% , 24% and 24.5 % respectively

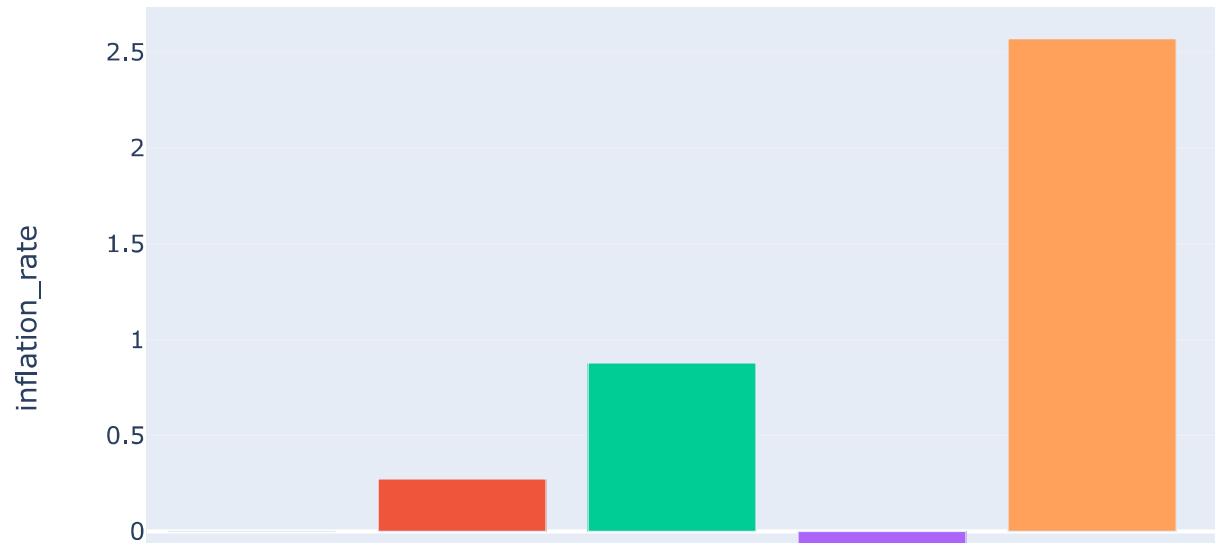
While South Africa and Rwanda have the lowest Investment_percent_GDP of about 17% and 19% respectively over the past 23 years

```
In [ ]:
```

```
In [73]: line_plot(G7_Inflation, 'country', 'inflation_rate' )
```



```
In [74]: ## fill the missing value with zero  
px.bar(G7_Inflation.fillna(0), x='country', y='inflation_rate', animation_frame=
```

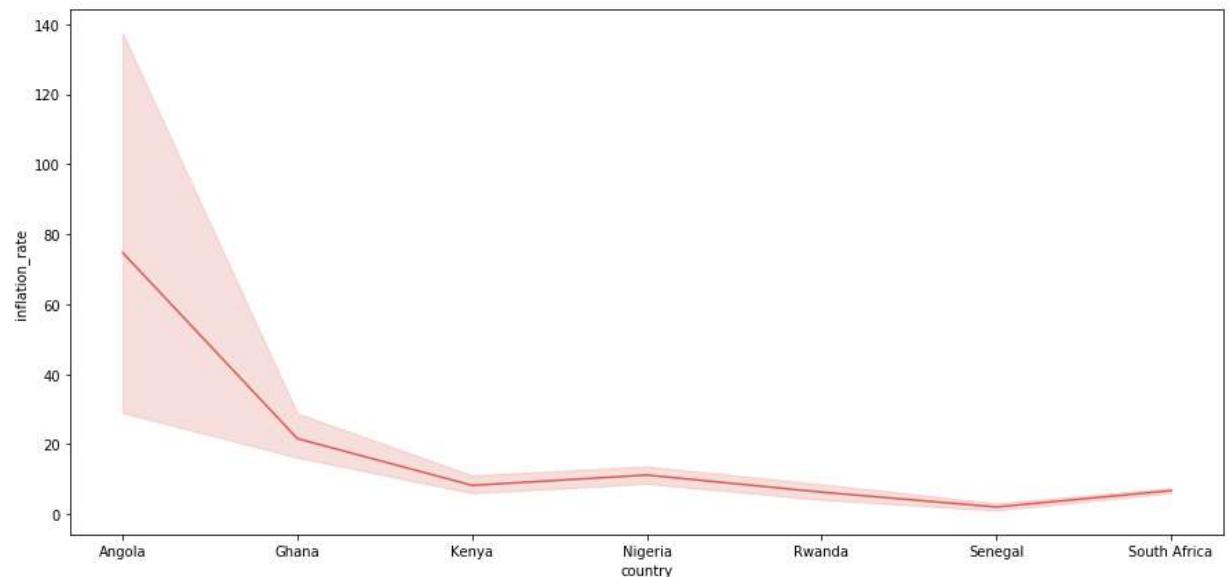


United Kingdom and United States have the highest average inflation_rate in the last 23 years amongst G7 countries of 1.8% and 1.7% respectively

while Japan and Germany had the lowest average inflation_rate of about 0.2% and 1.2% respectively within the time frame

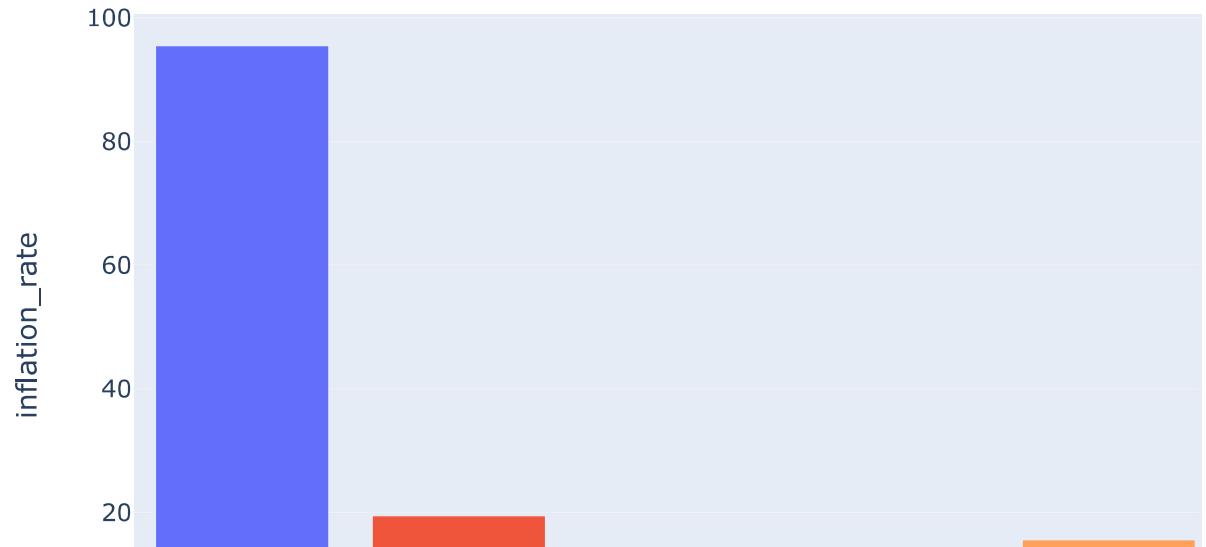
```
In [ ]:
```

```
In [75]: line_plot(SS7_Inflation, 'country', 'inflation_rate')
```



In [76]: *## fill the missing value with zero*

```
px.bar(SS7_Inflation, x='country', y='inflation_rate', animation_frame='years', c
```

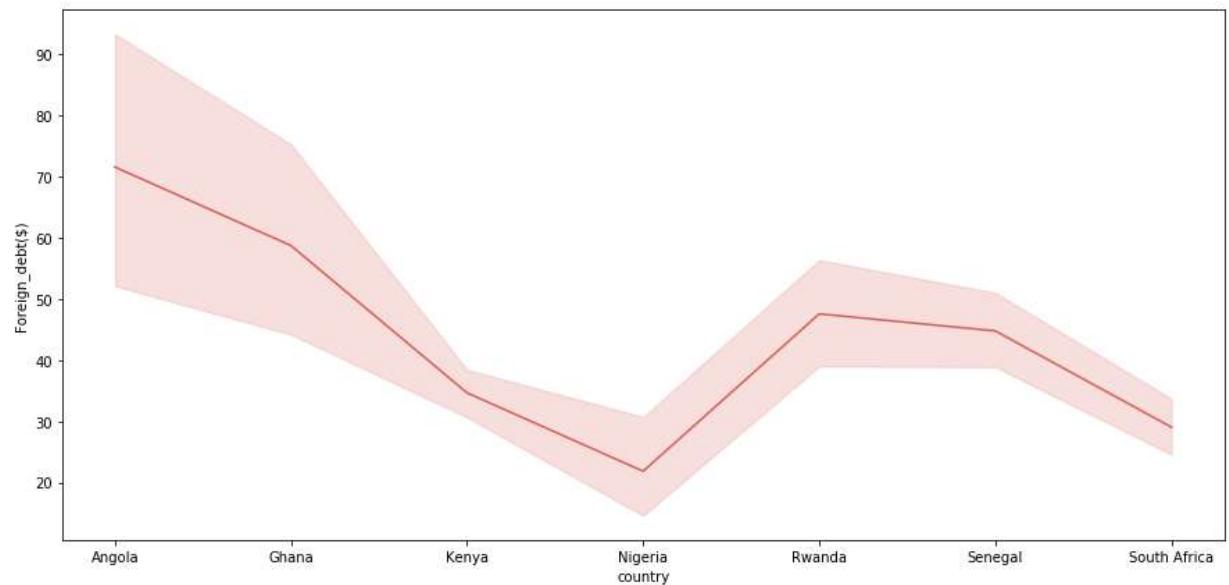


Angola and Ghana amongst the largest sub saharan economies have the highest inflation_rate of about 75% and 20% respectively

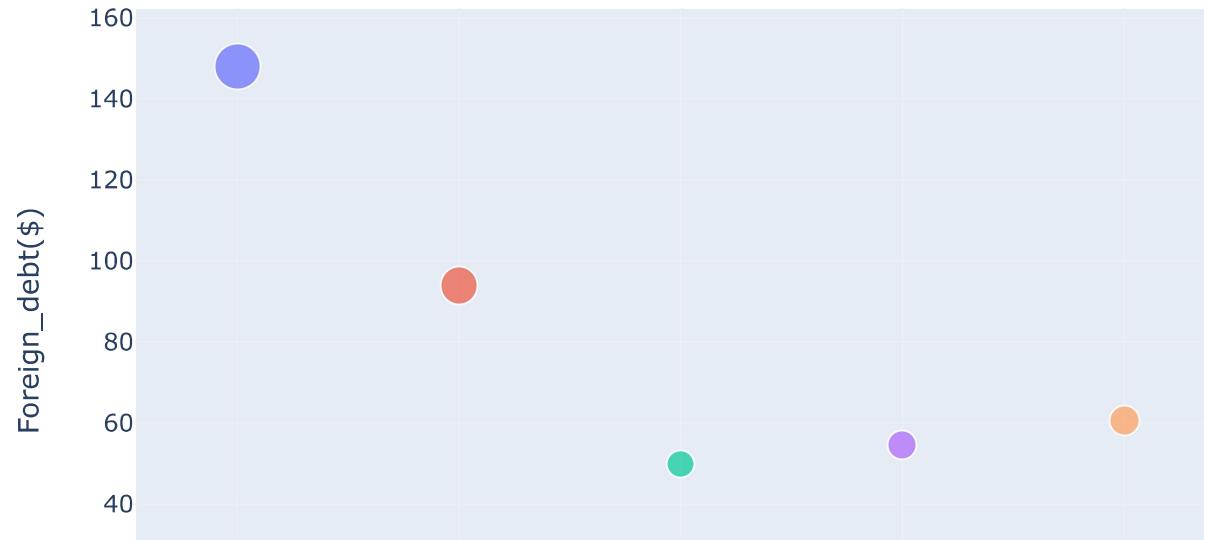
while Senegal and South Africa have the lowest average inflation_rate from 1997 to 2020 of about 2.5% and 6% respectively

In []:

```
In [77]: line_plot(Foreign_debt_SS7, 'country', 'Foreign_debt($)')
```



```
In [78]: px.scatter(Foreign_debt_SS7, x='country', y="Foreign_debt($)", animation_frame='y'
size= 'Foreign_debt($)')
```

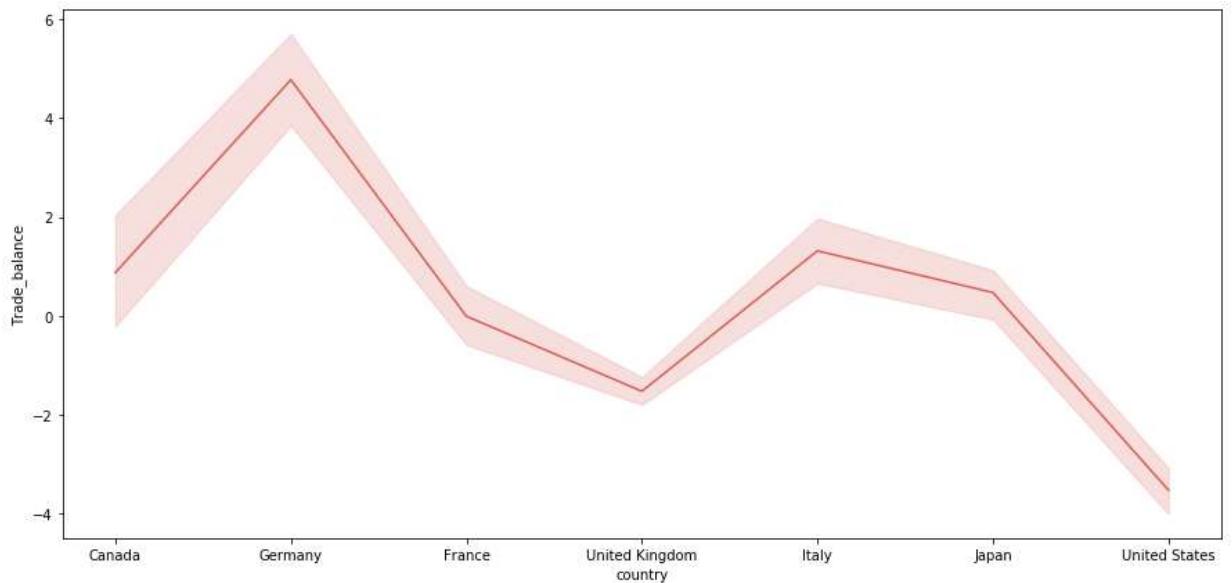


Angola and Ghana from 1997 to 2020 have accumulated the highest average foreign debt of about 72billion and \$60billion respectively

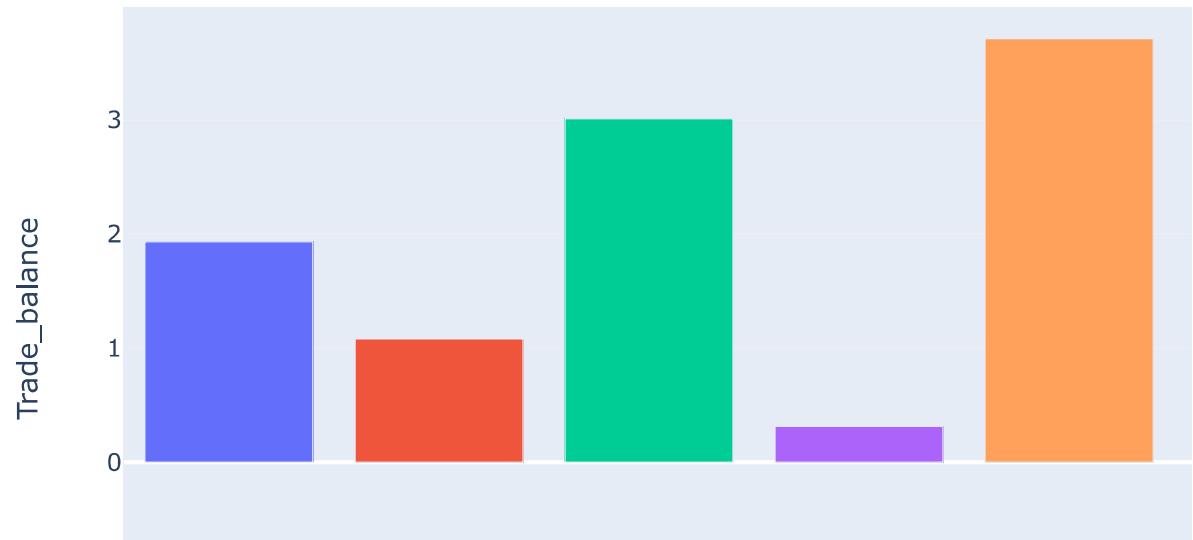
while Nigeria and South Africa have the lowest foreign debt of about 25billion and \$28 billion respectively

```
In [ ]:
```

```
In [79]: line_plot(G7_Trade_balance, 'country', 'Trade_balance')
```



```
In [80]: ## fill the missing value with zero  
px.bar(G7_Trade_balance, x='country', y='Trade_balance', animation_frame='years')
```



```
In [81]: G7_Trade_balance.query('country == "Italy" ').describe()
```

Out[81]:

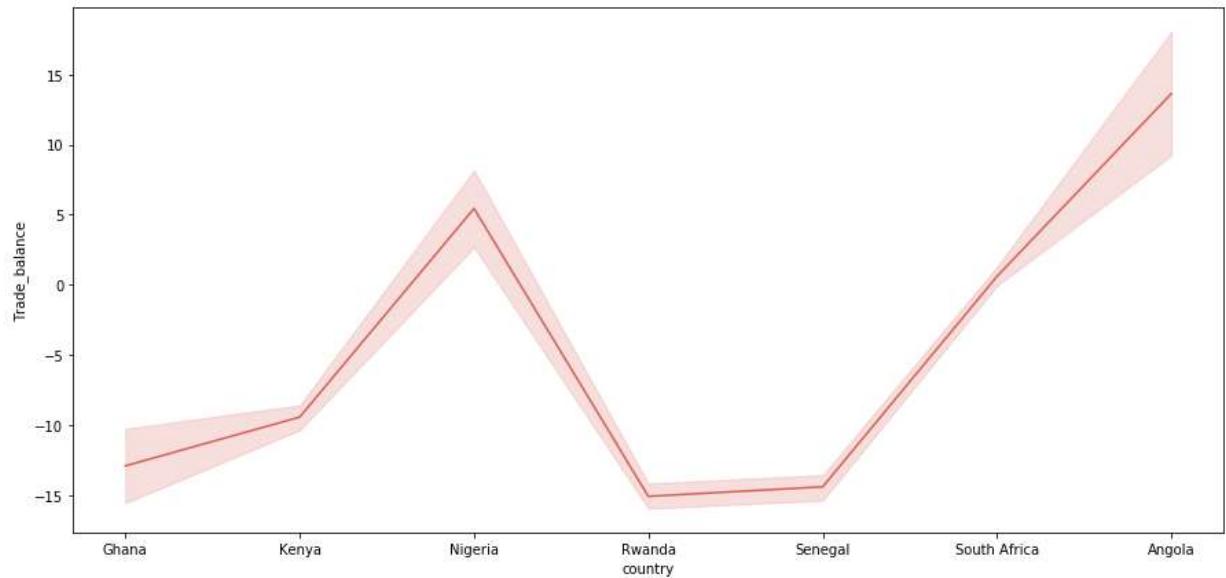
Trade_balance	
count	24.000000
mean	1.311925
std	1.734696
min	-1.860000
25%	-0.158850
50%	1.180000
75%	2.927500
max	3.710000

Germany, Canada and Italy have the highest average trade_balance % of GDP of about 5%, 1% and 1.5% respectively amongst G7 countries

while the United States and United kingdom fom 1997 to 2020 have the lowest average trade_balance % of GDP of -3.8% and 1.8% respectively

In []:

In [82]: `line_plot(SS7_Trade_balance, 'country', 'Trade_balance')`



```
In [83]: ## fill the missing value with zero  
px.bar(SS7_Trade_balance, x='country', y='Trade_balance', animation_frame='years')
```



```
In [84]: SS7_Trade_balance.query('country == "Rwanda"').describe()
```

Out[84]:

Trade_balance	
count	24.000000
mean	-15.093333
std	2.364047
min	-18.700000
25%	-16.525000
50%	-15.500000
75%	-13.750000
max	-8.640000

Angola and Nigeria amongst the top 7 Sub saharan economy had the highest trade_balance % of the GDP with about 13.6% and 5.8% respectively

Rwanda and Senegal have the lowest average trade_balance % of GDP of about -15% and 14% respectively

In []:

create a 2 new dataframe that contains trade_balance, inflation_rate and investment_percent_GDP of SS7 and G7 countries

In [85]: `from functools import reduce`

In [86]: `dataframe1=[Foreign_debt_SS7, SS7_Trade_balance, SS7_Inflation, Investment_ss7]`

In [87]: `SS7_economy= reduce(lambda left, right:pd.merge(left, right, on=['years', 'country'], how='outer'), dataframe1)`

In [88]: `SS7_economy`

Out[88]:

	country	years	Foreign_debt(\$)	Trade_balance	inflation_rate	Investment_percent_GDP
0	Angola	1997	148.0	NaN	95.50	NaN
1	Ghana	1997	93.9	-20.60	19.50	24.8
2	Kenya	1997	49.9	-8.68	11.40	15.1
3	Nigeria	1997	54.5	5.84	5.06	38.5
4	Rwanda	1997	60.6	-17.90	15.60	13.8
...
163	Kenya	2020	38.5	-8.00	5.18	19.7
164	Nigeria	2020	16.9	-7.74	7.85	29.4
165	Rwanda	2020	81.1	-16.30	8.27	24.5
166	Senegal	2020	71.7	-14.90	2.34	30.5
167	South Africa	2020	51.8	4.43	5.27	12.7

168 rows × 6 columns

In [89]: `dataframe2=[Investment_G7, G7_Trade_balance, G7_Inflation]`

In [90]: `G7_economy= reduce(lambda left, right:pd.merge(left, right, on=['years', 'country'], how='outer'), dataframe2)`

In [91]: G7_economy

Out[91]:

	country	years	Investment_percent_GDP	Trade_balance	inflation_rate
0	Canada	1997	21.2	1.9300	NaN
1	Germany	1997	23.4	1.0800	0.273
2	France	1997	19.5	3.0100	0.878
3	United Kingdom	1997	17.6	0.3160	-0.349
4	Italy	1997	19.5	3.7100	2.570
...
163	France	2020	23.8	-2.0200	2.520
164	United Kingdom	2020	16.7	0.1330	5.350
165	Italy	2020	17.6	3.6900	1.170
166	Japan	2020	25.5	-0.0168	0.842
167	United States	2020	21.1	-3.1100	1.210

168 rows × 5 columns

In [92]: SS7_economy.describe()

Out[92]:

	Foreign_debt(\$)	Trade_balance	inflation_rate	Investment_percent_GDP
count	168.000000	165.000000	168.000000	165.000000
mean	43.978036	-4.930888	18.656512	21.784242
std	33.156632	11.621256	56.472560	5.914327
min	4.950000	-25.300000	-16.800000	12.000000
25%	22.800000	-14.300000	4.290000	16.800000
50%	35.750000	-8.400000	7.720000	21.000000
75%	54.950000	1.160000	15.200000	25.600000
max	223.000000	32.300000	558.000000	42.800000

In [93]: `G7_economy.describe()`

Out[93]:

	Investment_percent_GDP	Trade_balance	inflation_rate
count	168.000000	168.000000	167.000000
mean	21.514881	0.336796	1.329031
std	2.881505	2.955122	1.240221
min	14.900000	-5.690000	-2.310000
25%	19.700000	-1.862500	0.758000
50%	21.400000	-0.165900	1.500000
75%	23.425000	1.982500	2.030000
max	31.100000	7.600000	5.350000

In [94]: `fig, axes= plt.subplots(1,1, sharex=True, sharey=False)`

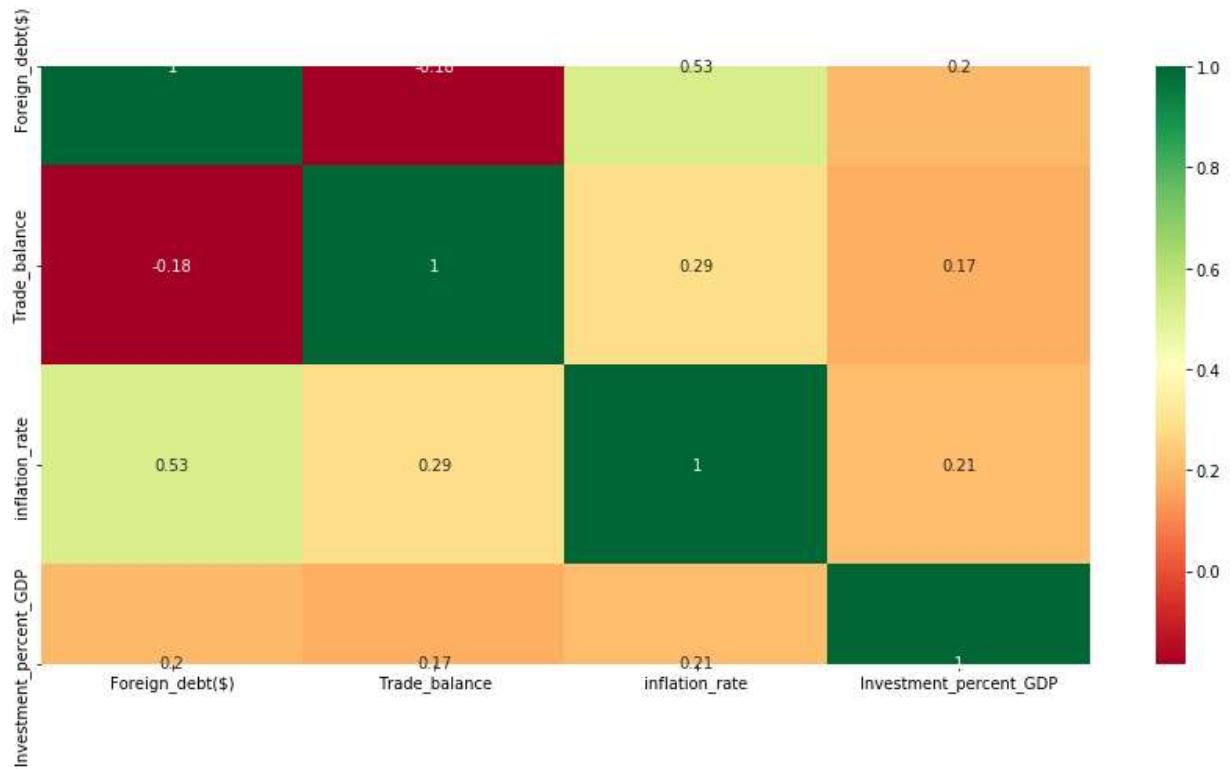
`fig.set_figheight(7)`

`fig.set_figwidth(15)`

`fig.set_facecolor('white')`

`sns.heatmap(SS7_economy.corr(), annot = True, cmap = "RdYlGn")`

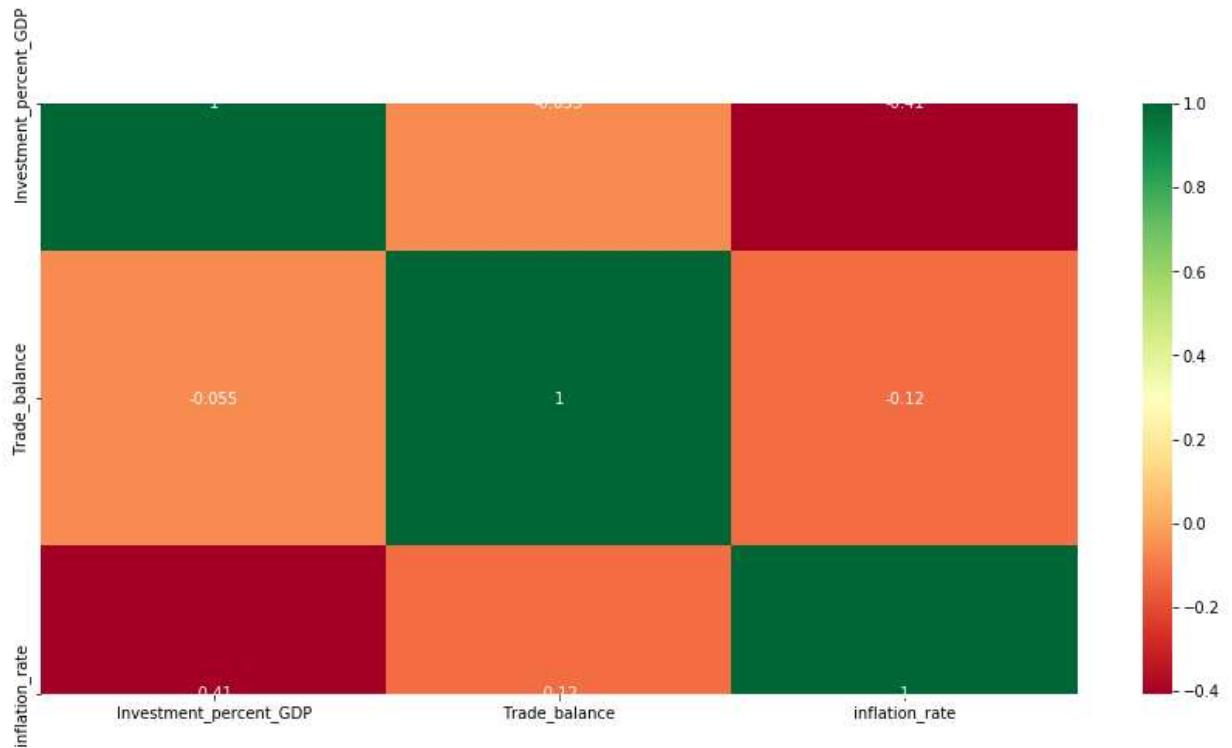
`plt.show()`



```
In [95]: fig, axes= plt.subplots(1,1, sharex=True, sharey=False)
fig.set_figheight(7)
fig.set_figwidth(15)
fig.set_facecolor('white')

sns.heatmap(G7_economy.corr(), annot = True, cmap = "RdYlGn")

plt.show()
```



Research Question 2 (Replace this header name!)

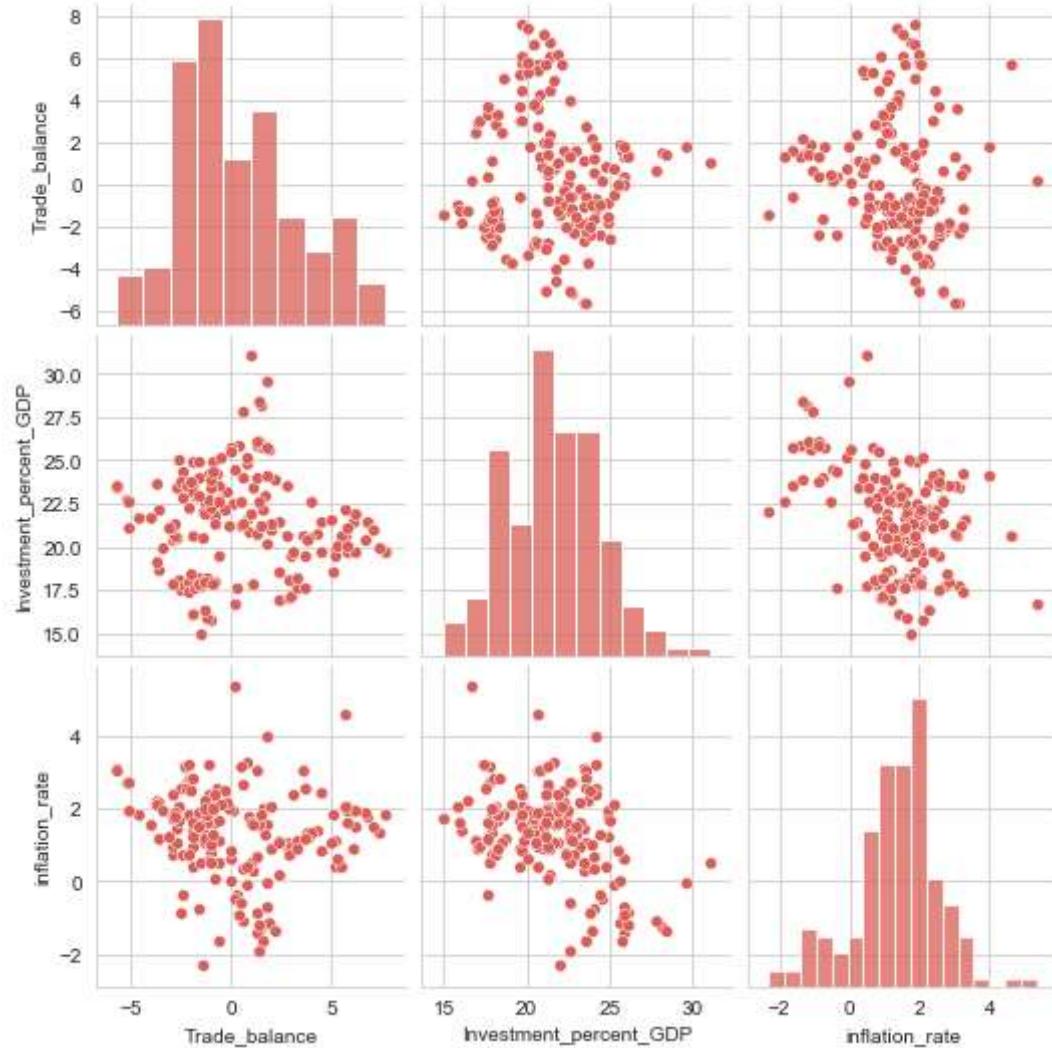
```
In [96]: # Continue to explore the data to address your additional research
#   questions. Add more headers as needed if you have more questions to
#   investigate.
sns.set_style(style= 'whitegrid')
```

```
In [97]: numeric_cols1=['Trade_balance', 'Investment_percent_GDP', 'inflation_rate']
numeric_cols2= ['Foreign_debt($)', 'Trade_balance', 'Investment_percent_GDP', 'ir]
```

In [98]:

```
sns.pairplot(data=G7_economy, vars=numeric_cols1)
```

Out[98]: <seaborn.axisgrid.PairGrid at 0x219cfb90e48>

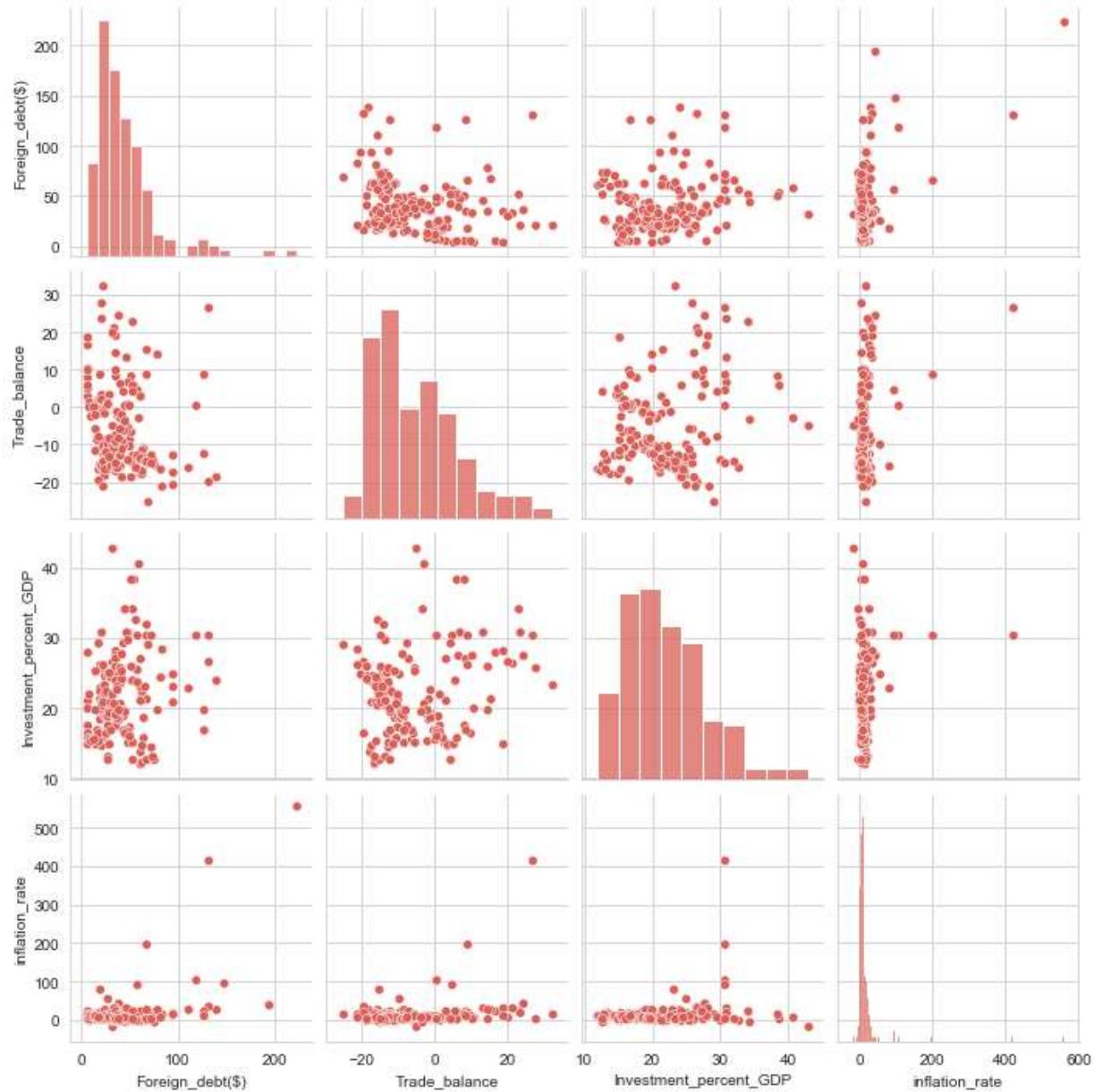


In []:

In [99]:

```
sns.pairplot(data=SS7_economy, vars=numeric_cols2)
```

Out[99]: <seaborn.axisgrid.PairGrid at 0x219d0cbb2c8>

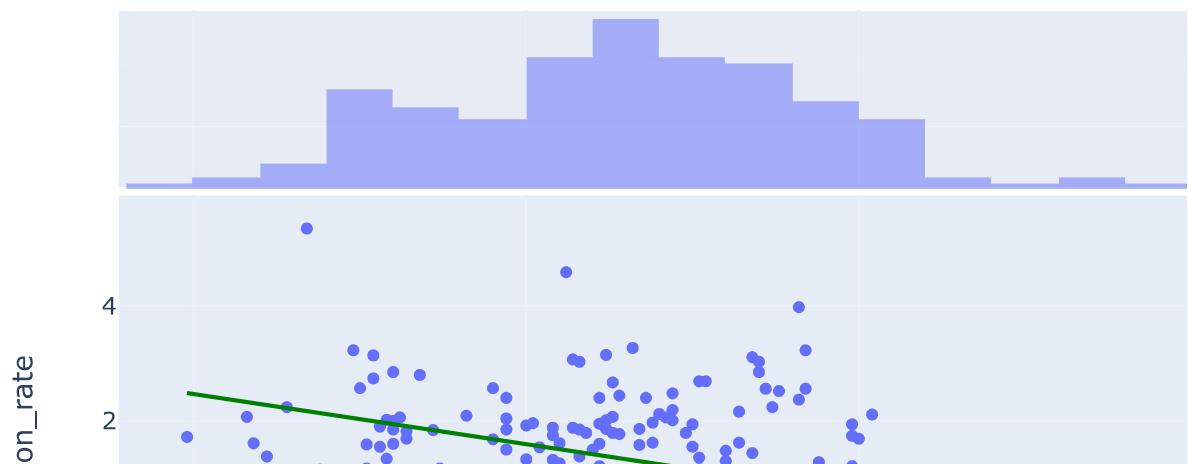


```
In [100]: corr_value= G7_economy['Investment_percent_GDP'].corr(G7_economy['inflation_rate'])
fig=px.scatter(
    data_frame=G7_economy,
    x='Investment_percent_GDP',
    y='inflation_rate',
    marginal_x='histogram',
    marginal_y='histogram',
    trendline='ols',
    trendline_color_override='green',
    title= f"Correlation between inflation_rate and Investment_percent_GDP is:{corr_value}"
)
fig.show()
```

C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\tools_testing.py:19: FutureWarning:

pandas.util.testing is deprecated. Use the functions in the public API at pandas.testing instead.

Correlation between inflation_rate and Investment_percent_GDP



we can see that Investment percent GDP and inflation rate has a general negative

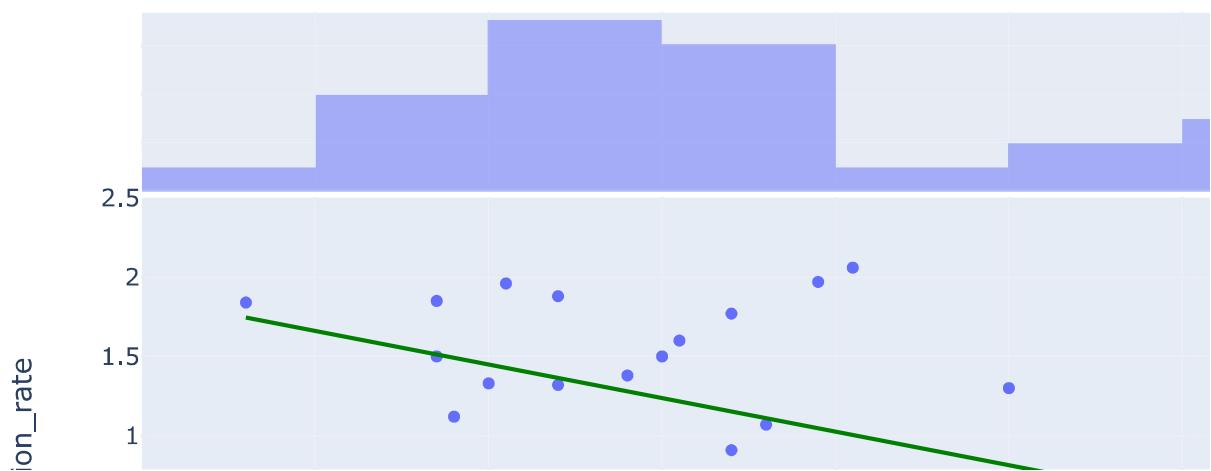
relationship between G7 countries

but for individual countries it might be different

In []:

```
In [101]: corr_value= G7_economy['Investment_percent_GDP'].corr(G7_economy['inflation_rate'])
fig=px.scatter(
    data_frame=G7_economy.query('country == "Germany"'),
    x='Investment_percent_GDP',
    y='inflation_rate',
    marginal_x='histogram',
    marginal_y='histogram',
    trendline='ols',
    trendline_color_override='green',
    title= f"Correlation between inflation_rate and Investment_percent_GDP is:{corr_value}"
)
fig.show()
```

Correlation between inflation_rate and Investment_percent_GDP



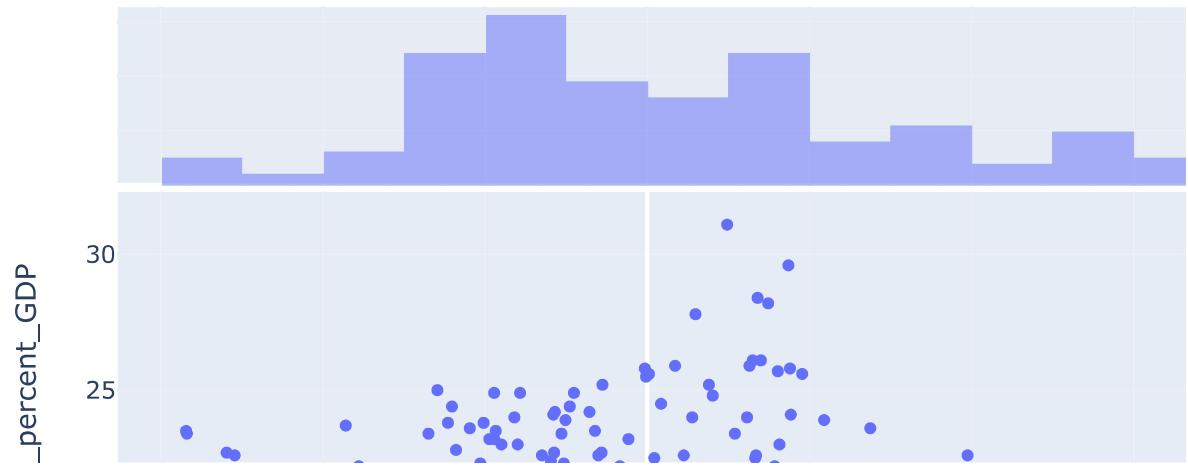
In [169]: *### a function for correlation plot between variables*

```
def corr_plot(data, variable1:str, variable2:str):
    """
    data= dataframe,
    variable1= column_x,
    variable2= column_y

    """
    x=variable1
    y=variable2
    corr_value1= data[variable1].corr(data[variable2])
    fig=px.scatter(
        data_frame=data,
        x=variable1,
        y=variable2,
        marginal_x='histogram',
        marginal_y='histogram',
        trendline='ols',
        trendline_color_override='green',
        title= f"Correlation between {x} and {y} is:{corr_value1:2f}"
    )
    fig.show()
```

```
In [103]: corr_plot(G7_economy, 'Trade_balance', 'Investment_percent_GDP')
```

Correlation between Trade_balance and Investment_percent_GDP

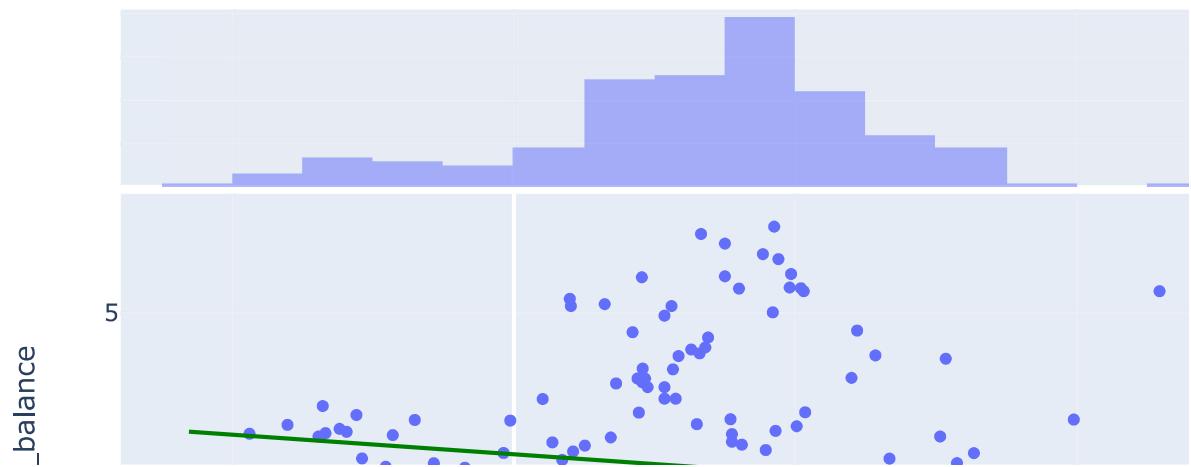


for G7 countries there seems to be no substantial relationship between investment_percent_GDP and Trade_balance % GDP

```
In [ ]:
```

```
In [104]: corr_plot(G7_economy, 'inflation_rate', 'Trade_balance')
```

Correlation between inflation_rate and Trade_balance is:-0.12341



```
In [105]: ##### there is a negative relationship between inflation_rate and investment_perce
```

```
In [ ]:
```

```
In [106]: corr_plot(SS7_economy, 'Foreign_debt($)', 'inflation_rate' )
```

Correlation between Foreign_debt(\$) and inflation_rate is: 0.5255

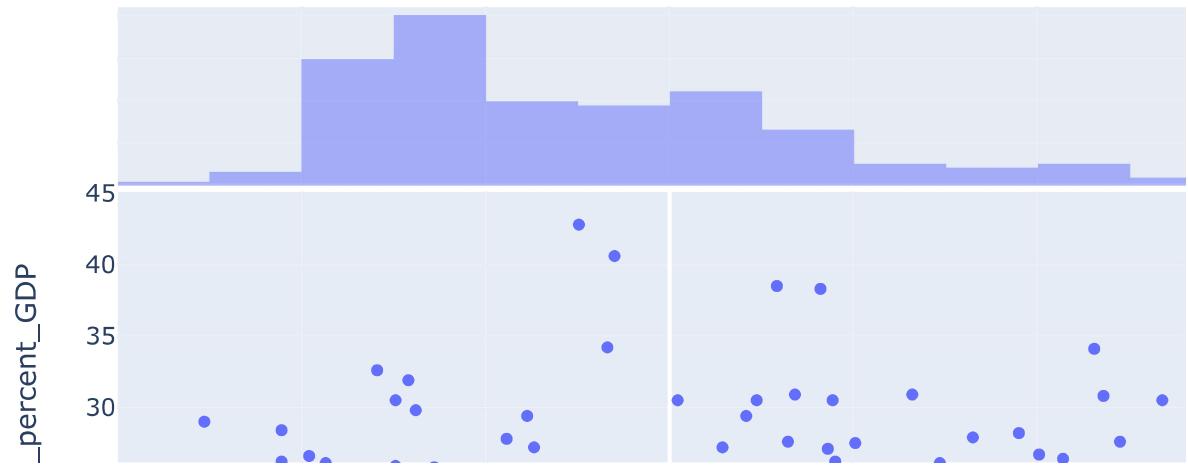


there is a positive relationship between inflation_rate and Foreign_debt(\$) amongst SS7 countries

```
In [ ]:
```

```
In [107]: corr_plot(SS7_economy, 'Trade_balance', 'Investment_percent_GDP')
```

Correlation between Trade_balance and Investment_percent_GDP

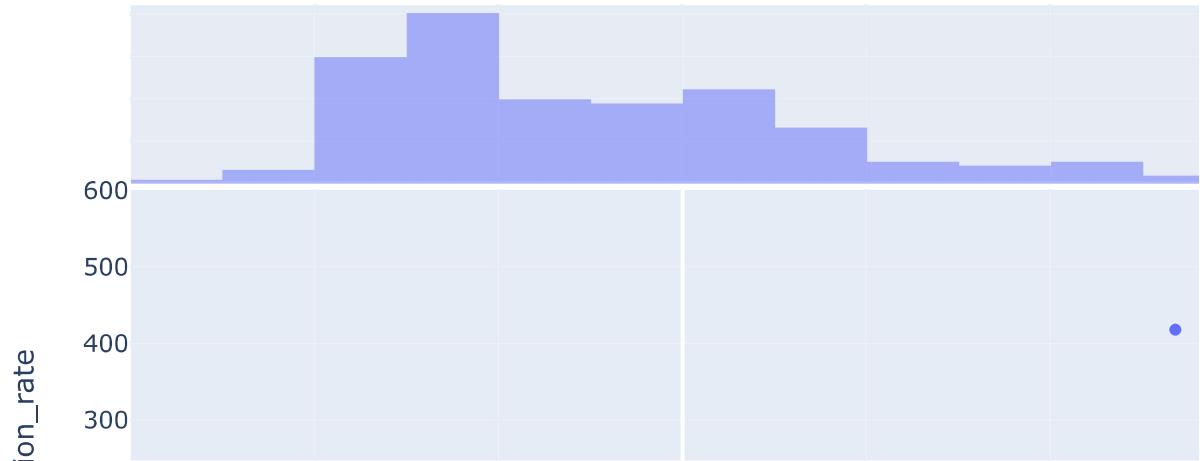


there is a very weak positive relationship between Investment_percent_GDP and Trade_balance among the top 7 sub saharan economies

```
In [ ]:
```

```
In [108]: corr_plot(SS7_economy, 'Trade_balance', 'inflation_rate')
```

Correlation between Trade_balance and inflation_rate is:0.290861

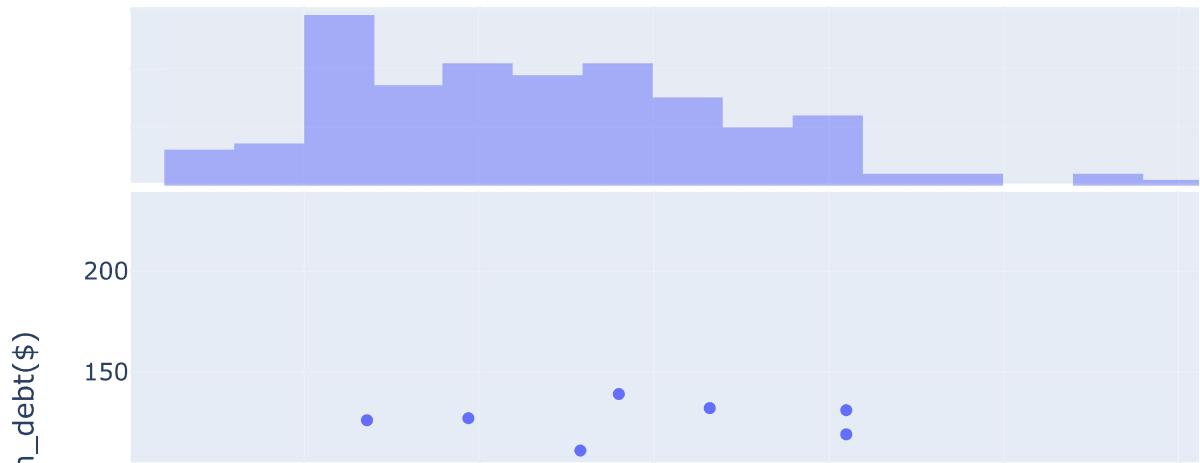


there is a positive relationship between inflation_rate and Trade_balance amongst SS7 countries

```
In [ ]:
```

```
In [109]: corr_plot(SS7_economy, 'Investment_percent_GDP', 'Foreign_debt($)')
```

Correlation between Investment_percent_GDP and Foreign_debt(

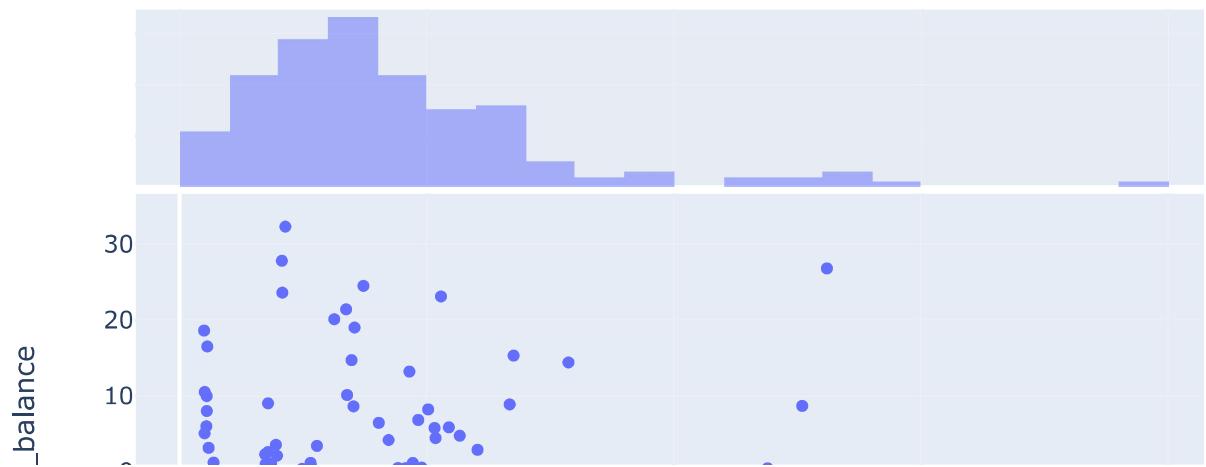


there is weak or no relationship between Investment_percent_GDP and Foreign_debt in the SS7_economy

```
In [ ]:
```

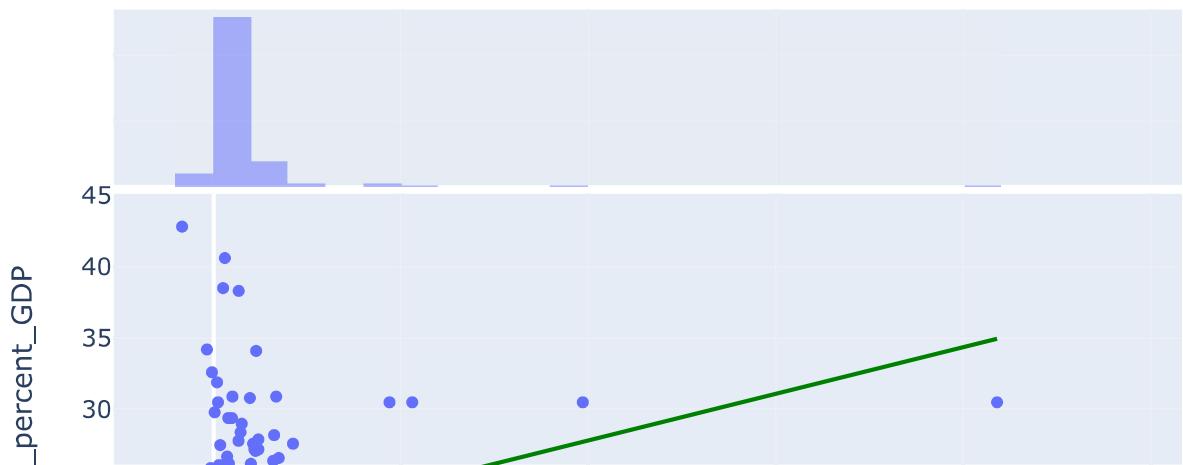
```
In [110]: corr_plot(SS7_economy, 'Foreign_debt($)', 'Trade_balance' )
```

Correlation between Foreign_debt(\$) and Trade_balance is:-0.18



```
In [111]: corr_plot(SS7_economy, 'inflation_rate', 'Investment_percent_GDP' )
```

Correlation between inflation_rate and Investment_percent_GDP



Research answer 2:

2a. Inflation have an effect on Foreign_debt . the higher the inflation_rate the higher the forien_debt

2b. From our analysis the relationship between Foreign_debt and Gross capital formation(investment_percent_GDP) is positive but it is very weak to conclude

2c. There isn't any substantial effect between Foreign_debt and Trade_balance

2d. Yes inflation_rate have an effect on Gross capital formation (investment_percent_GDP) ,but it is too weak

2e. Amongst G7 countries (Canada, Italy and Germany) have the highest trade_balance , while Nigeria and Angola have the highest trade_balance amongst the top 7 sub saharan economies

2f. The effect of Gross capital formation on Trade_balance is too weak

In []:

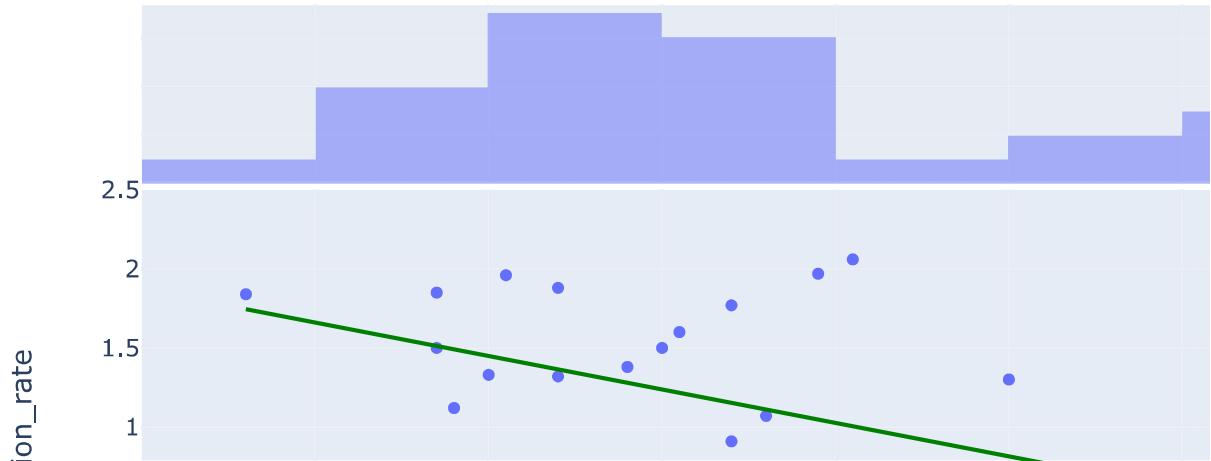
In [167]: *### a function for correlation plot between variables but specifies the country*

```
def corr_plot1(data, variable1:str, variable2:str, country_name: str):  
    """  
        data= dataframe,  
        variable1= column_x,  
        variable2= column_y,  
        country_name= name of country for specific analysis  
    """  
    x=variable1  
    y=variable2  
    z= country_name  
    new_data= data[data.country==country_name]  
    corr_value1= new_data[variable1].corr(new_data[variable2])  
    fig=px.scatter(  
        data_frame=new_data,  
        x=variable1,  
        y=variable2,  
        marginal_x='histogram',  
        marginal_y='histogram',  
        trendline='ols',  
        trendline_color_override='green',  
        title= f"Correlation between {x} and {y} of {z} is :{corr_value1:2f}"  
    )  
    fig.show()
```

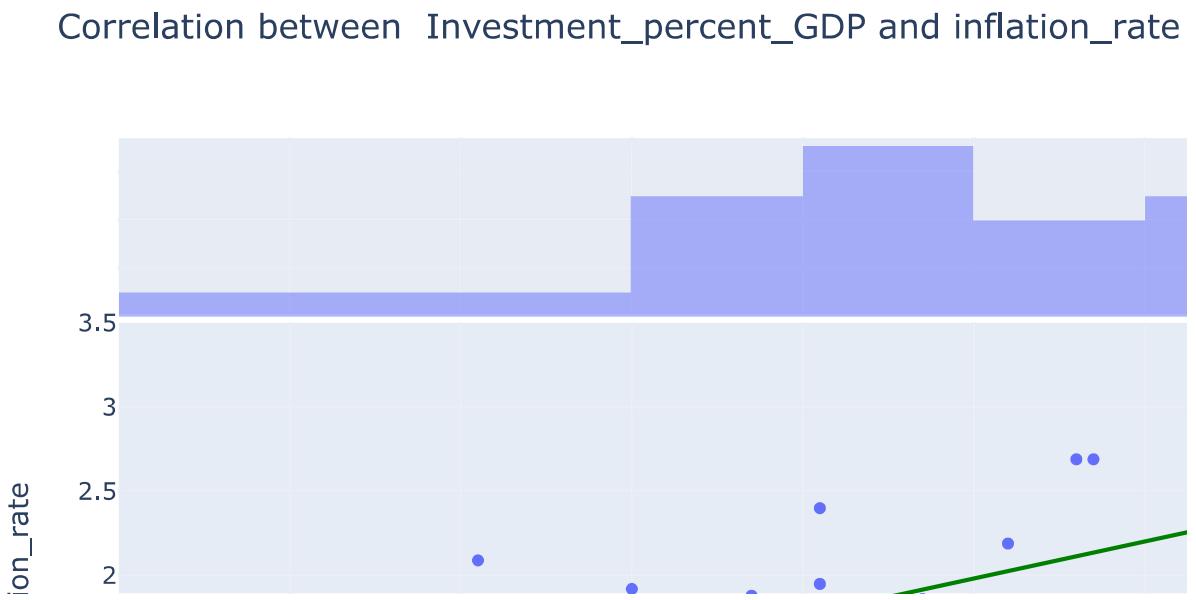
Inflation vs Investment (G7)

```
In [168]: corr_plot1(G7_economy, 'Investment_percent_GDP', 'inflation_rate', 'Germany')
```

Correlation between Investment_percent_GDP and inflation_rate



```
In [114]: corr_plot1(G7_economy, 'Investment_percent_GDP', 'inflation_rate', 'United States')
```



```
In [115]: corr_plot1(G7_economy, 'Investment_percent_GDP', 'inflation_rate', 'United Kingc
```

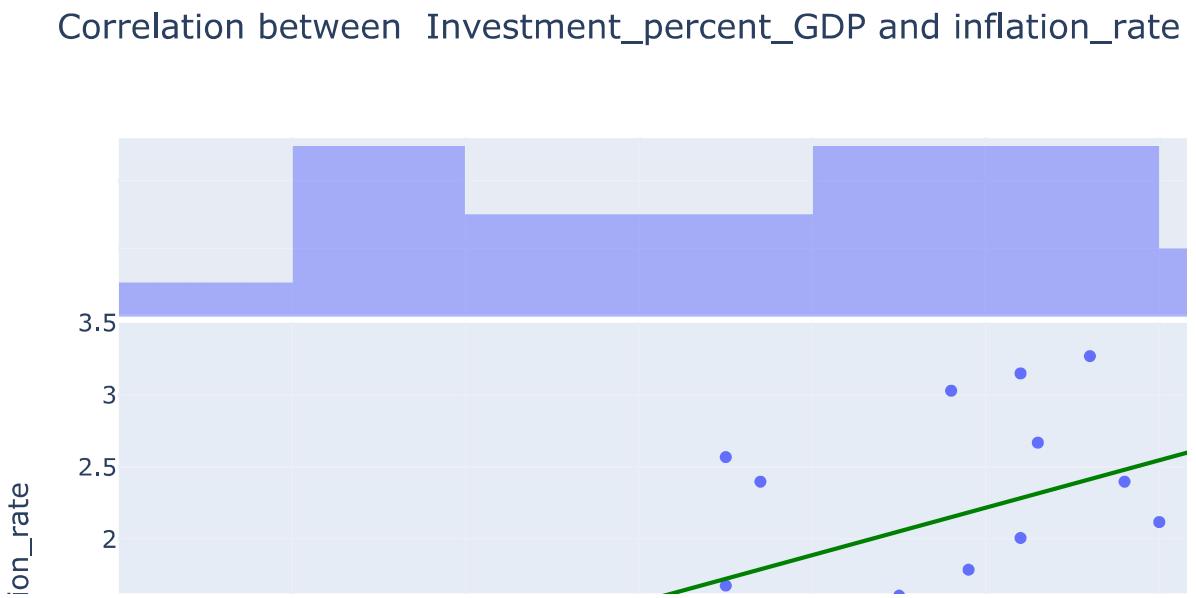


```
In [116]: corr_plot1(G7_economy, 'Investment_percent_GDP', 'inflation_rate', 'Japan')
```

Correlation between Investment_percent_GDP and inflation_rate

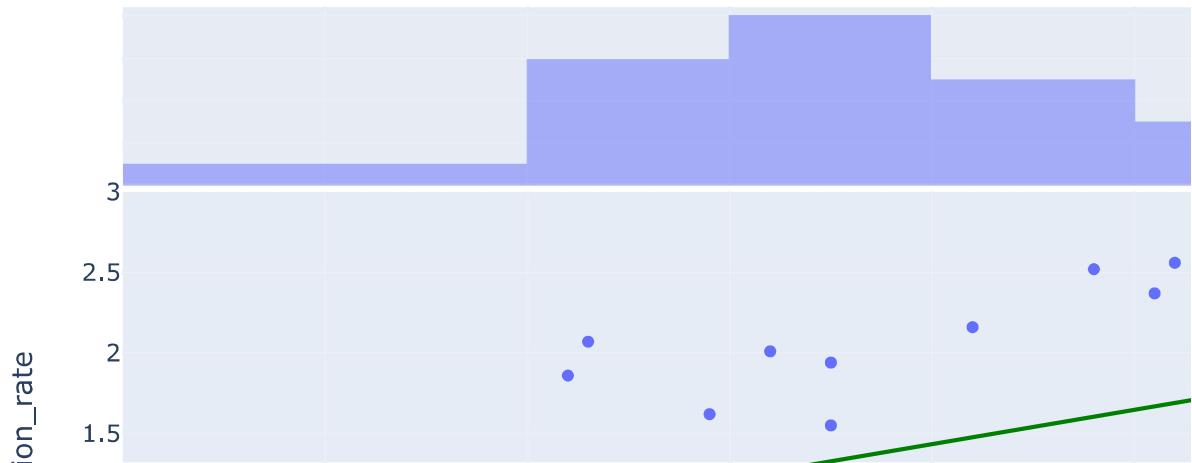


```
In [117]: corr_plot1(G7_economy, 'Investment_percent_GDP', 'inflation_rate', 'Italy')
```



```
In [118]: corr_plot1(G7_economy, 'Investment_percent_GDP', 'inflation_rate', 'France')
```

Correlation between Investment_percent_GDP and inflation_rate



As inflation_rate increases the investment_percent_GDP decreases for Germany

But reverse is the case for the United States, Italy and France

United Kingdom and Japan there is no relationship between Inflation_rate and Investment_percent_GDP

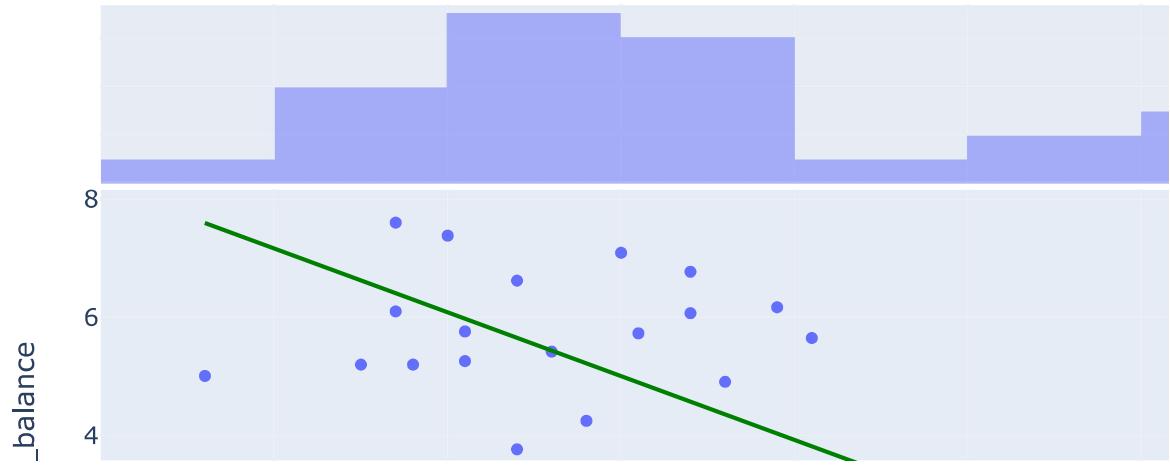
```
In [ ]:
```

```
In [ ]:
```

Trade_balance vs Investment (G7)

```
In [119]: corr_plot1(G7_economy, 'Investment_percent_GDP', 'Trade_balance', 'Germany')
```

Correlation between Investment_percent_GDP and Trade_balance

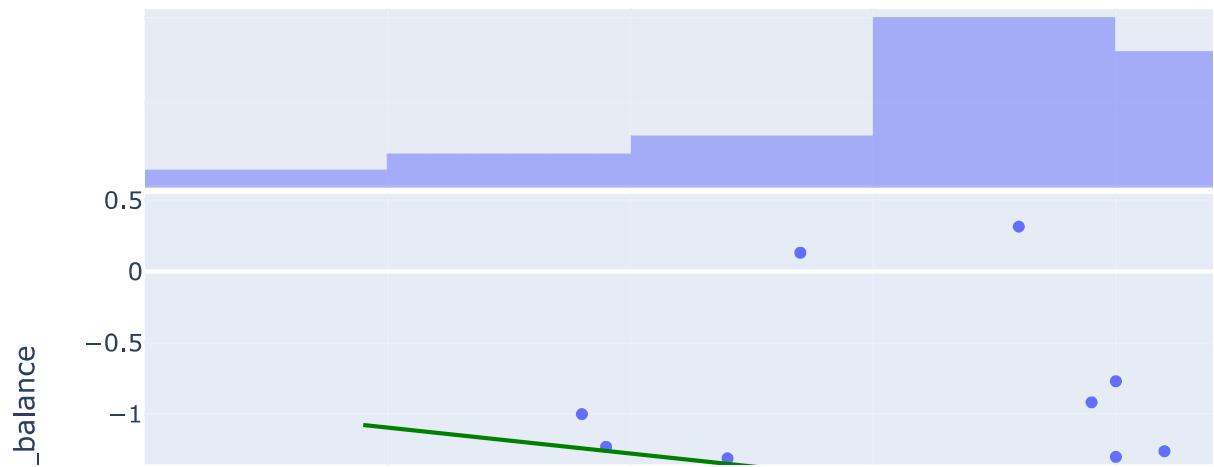


```
In [120]: corr_plot1(G7_economy, 'Investment_percent_GDP', 'Trade_balance', 'United States')
```



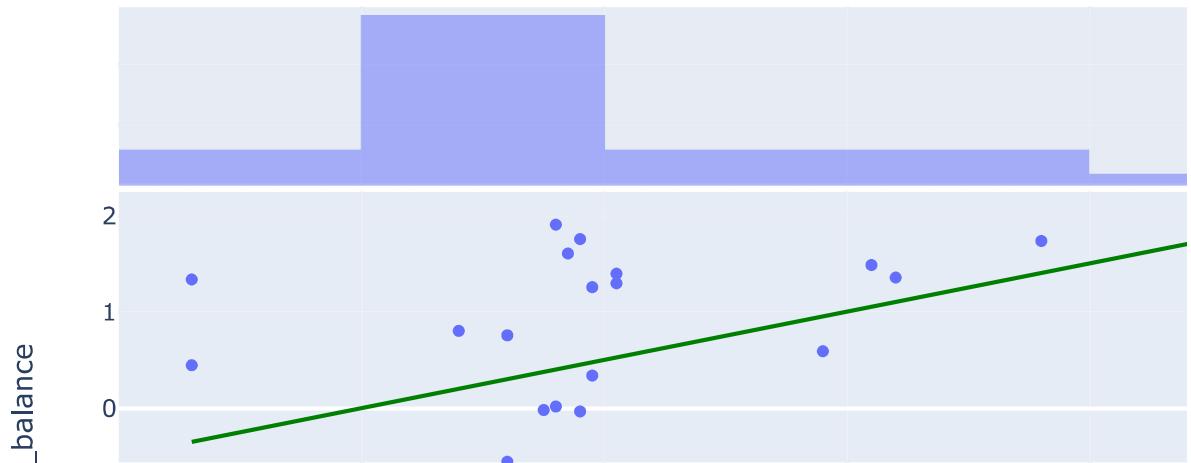
```
In [121]: corr_plot1(G7_economy, 'Investment_percent_GDP', 'Trade_balance', 'United Kingdom')
```

Correlation between Investment_percent_GDP and Trade_balance



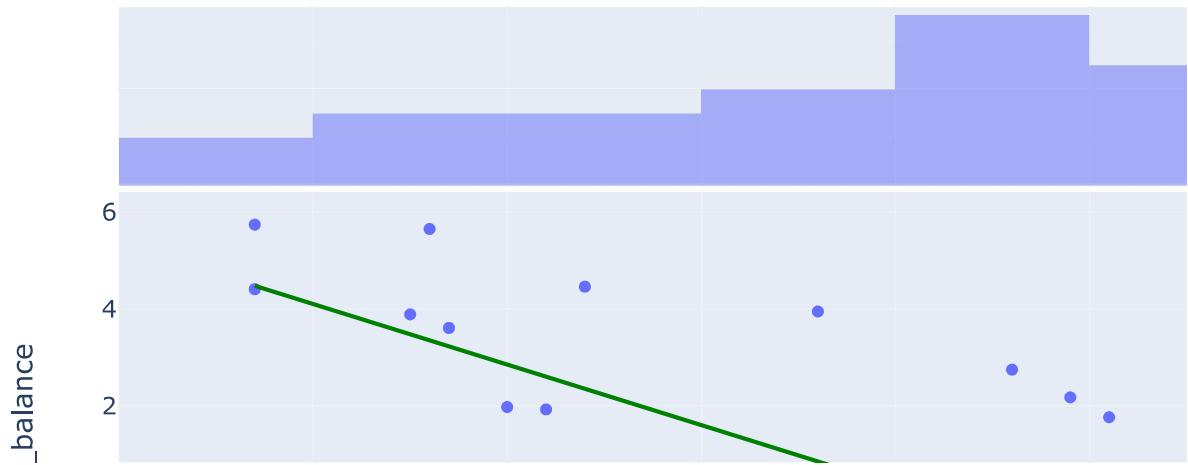
```
In [122]: corr_plot1(G7_economy, 'Investment_percent_GDP', 'Trade_balance', 'Japan')
```

Correlation between Investment_percent_GDP and Trade_balance



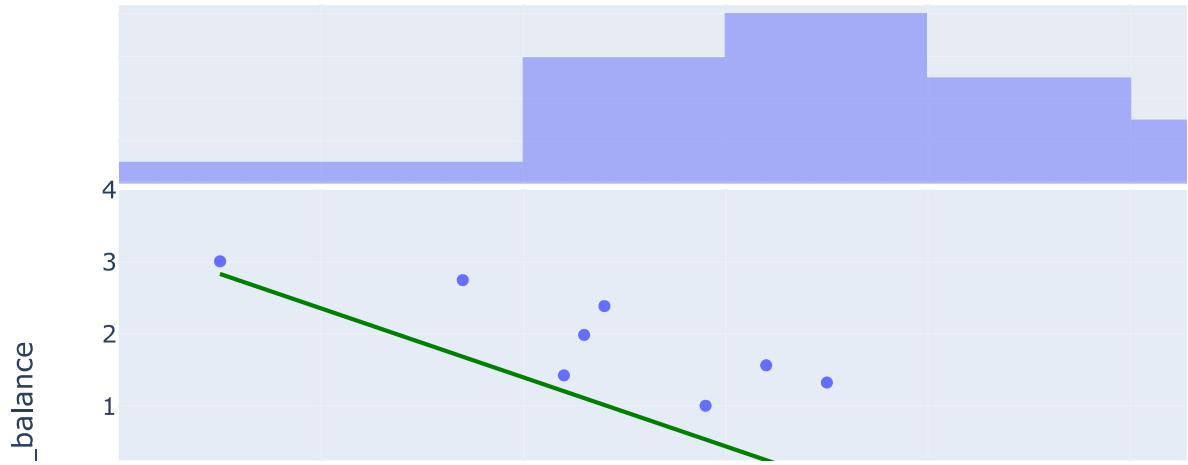
```
In [123]: corr_plot1(G7_economy, 'Investment_percent_GDP', 'Trade_balance', 'Canada')
```

Correlation between Investment_percent_GDP and Trade_balance



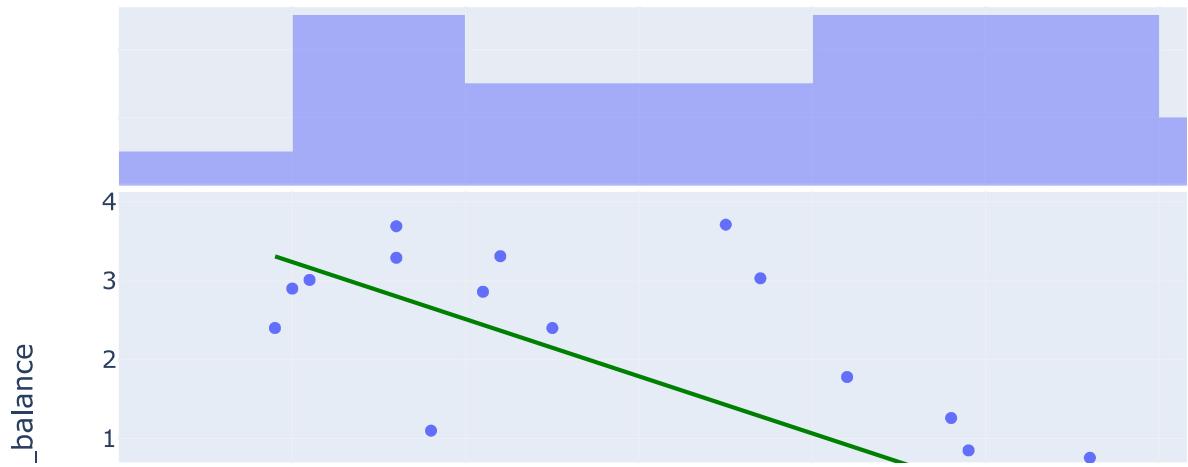
```
In [124]: corr_plot1(G7_economy, 'Investment_percent_GDP', 'Trade_balance', 'France')
```

Correlation between Investment_percent_GDP and Trade_balance



```
In [125]: corr_plot1(G7_economy, 'Investment_percent_GDP', 'Trade_balance', 'Italy')
```

Correlation between Investment_percent_GDP and Trade_balance



Germany, France , Italy and Canada show a very strong negative relationship between trade_balance and investment_percent_GDP

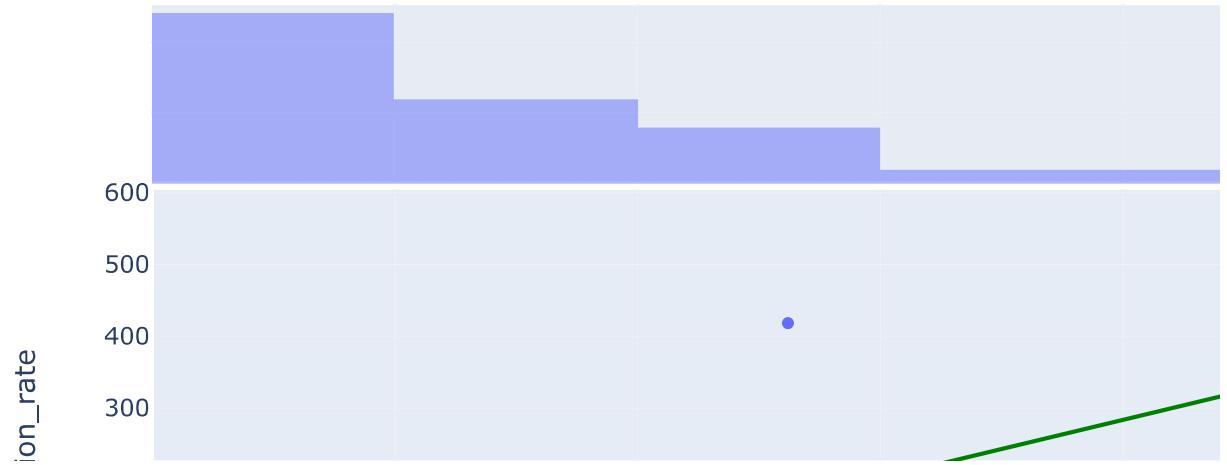
While Japan show a weak relationship between trade_balance and investment_percent_GDP

```
In [ ]:
```

Foreign_debt(\$) vs Inflation_rate

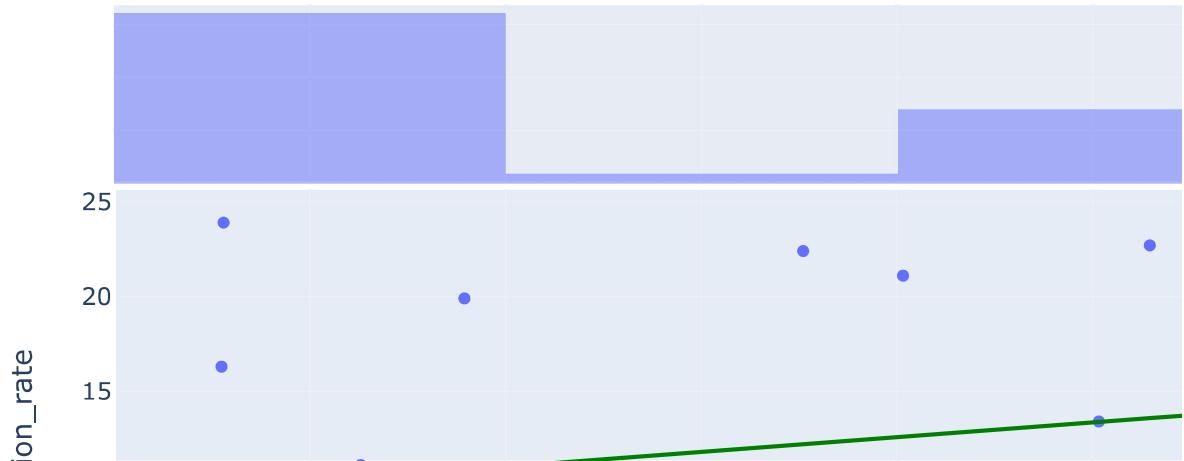
```
In [126]: corr_plot1(SS7_economy, 'Foreign_debt($)', 'inflation_rate', 'Angola')
```

Correlation between Foreign_debt(\$) and inflation_rate of Angola



```
In [127]: corr_plot1(SS7_economy, 'Foreign_debt($)', 'inflation_rate', 'Nigeria')
```

Correlation between Foreign_debt(\$) and inflation_rate of Nigeria

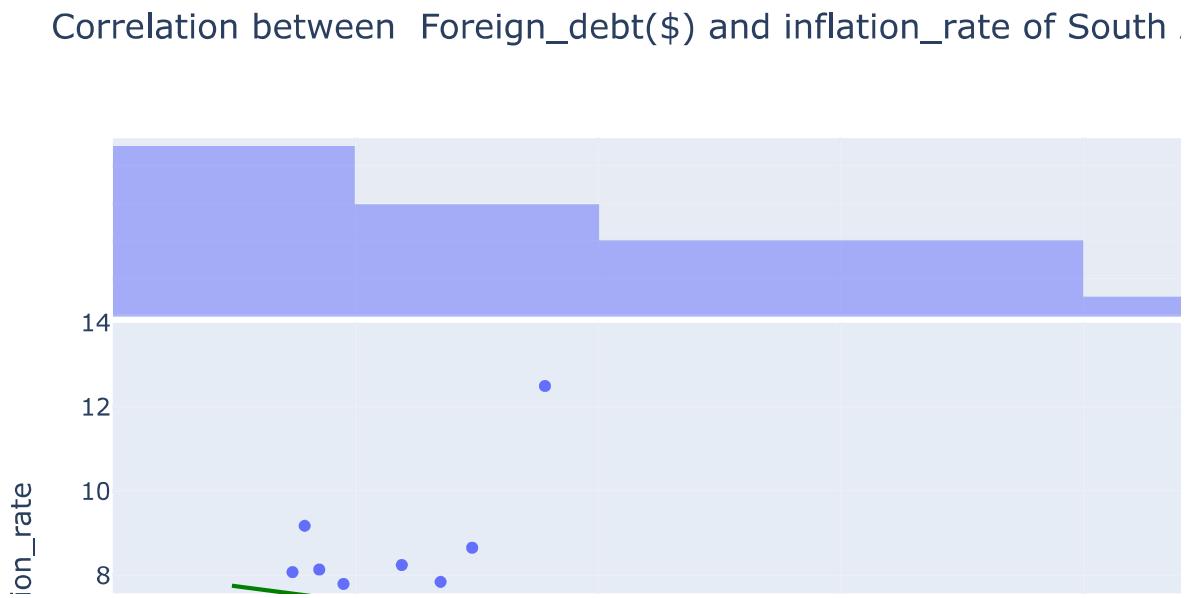


```
In [128]: corr_plot1(SS7_economy, 'Foreign_debt($)', 'inflation_rate', 'Rwanda')
```

Correlation between Foreign_debt(\$) and inflation_rate of Rwanda

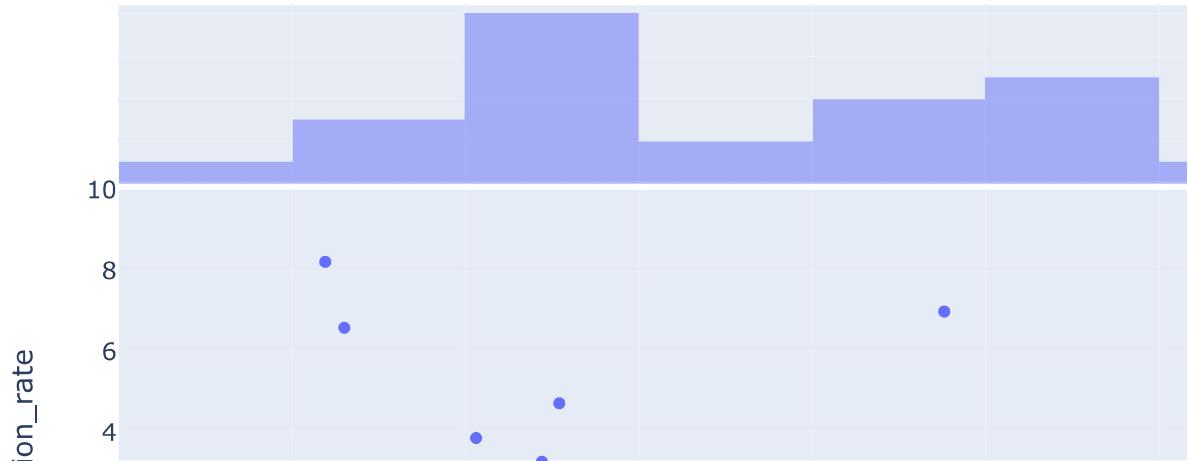


```
In [129]: corr_plot1(SS7_economy, 'Foreign_debt($)', 'inflation_rate', 'South Africa')
```



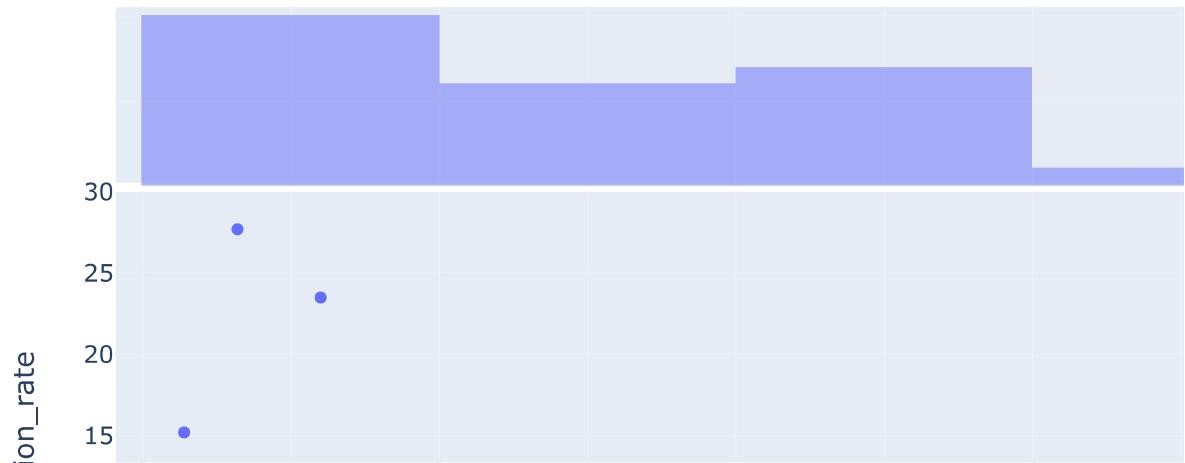
```
In [130]: corr_plot1(SS7_economy, 'Foreign_debt($)', 'inflation_rate', 'Senegal')
```

Correlation between Foreign_debt(\$) and inflation_rate of Senegal



```
In [131]: corr_plot1(SS7_economy, 'Foreign_debt($)', 'inflation_rate', 'Kenya')
```

Correlation between Foreign_debt(\$) and inflation_rate of Kenya



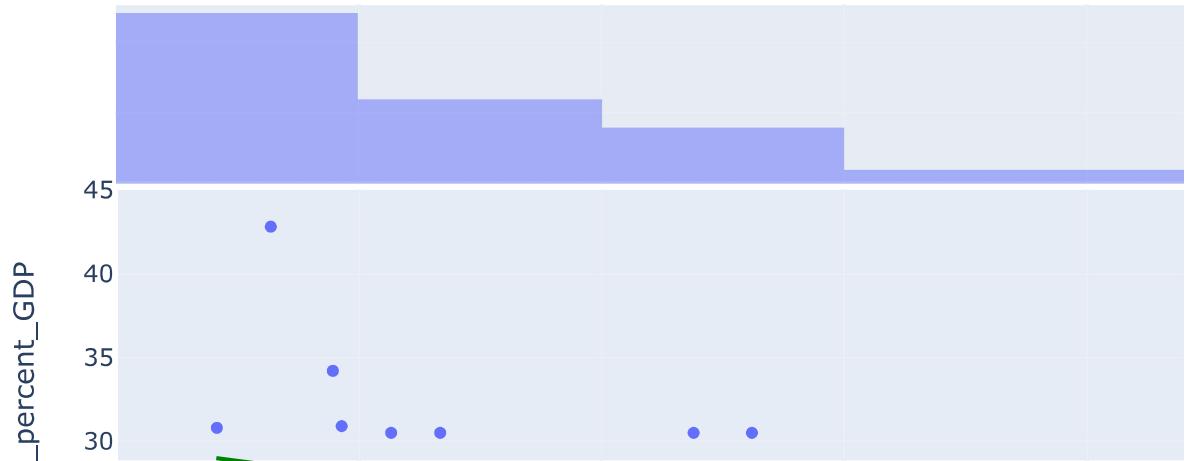
In Angola and Nigeria the higher the inflation_rate the higher the foreign_debt(\$)

While South Africa , Ghana, Kenya, Rwanda and senegal show a weak negative relationship between inflation_rate and Foreign_debt(\$)

```
In [ ]:
```

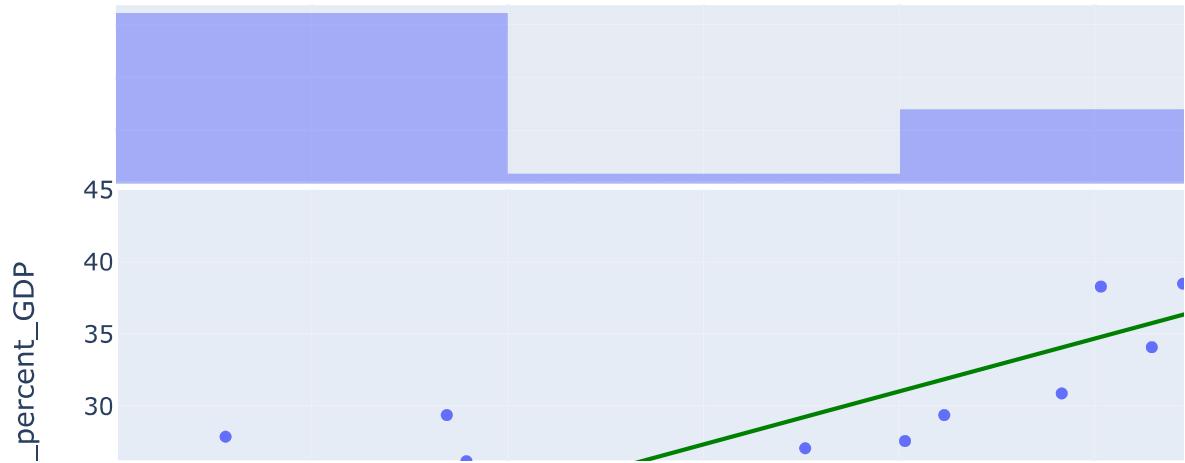
```
In [132]: corr_plot1(SS7_economy, 'Foreign_debt($)', 'Investment_percent_GDP', 'Angola')
```

Correlation between Foreign_debt(\$) and Investment_percent_GDP



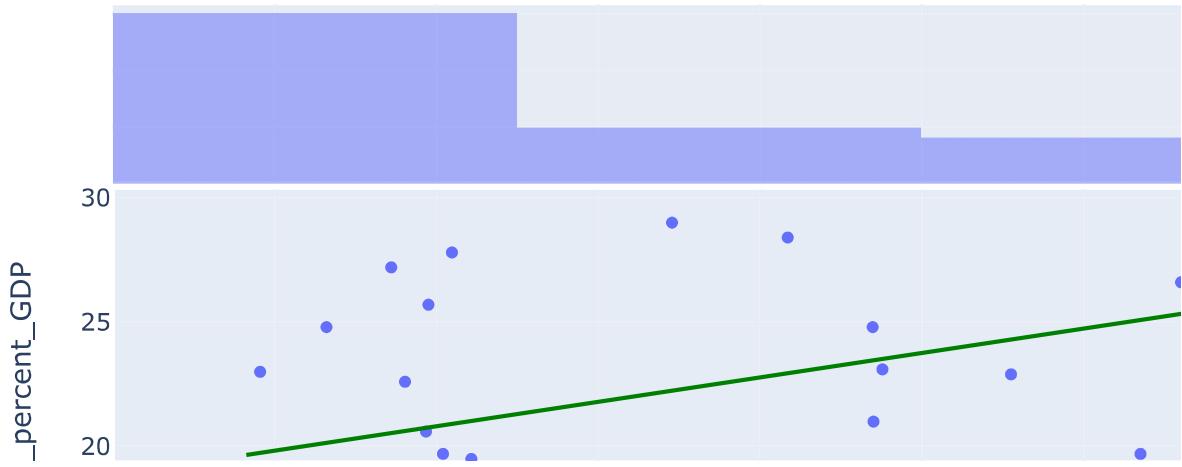
```
In [133]: corr_plot1(SS7_economy, 'Foreign_debt($)', 'Investment_percent_GDP', 'Nigeria')
```

Correlation between Foreign_debt(\$) and Investment_percent_GDP



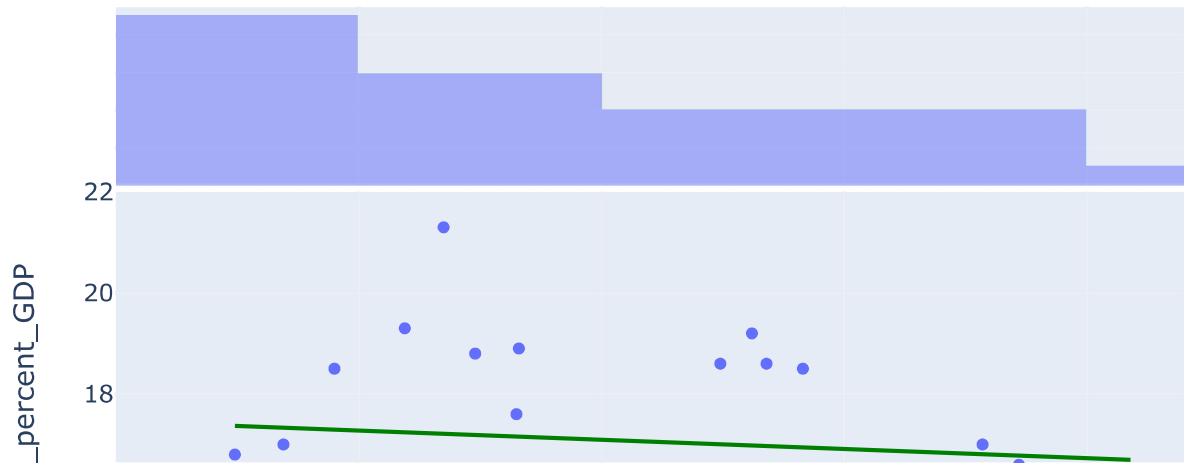
```
In [134]: corr_plot1(SS7_economy, 'Foreign_debt($)', 'Investment_percent_GDP', 'Ghana')
```

Correlation between Foreign_debt(\$) and Investment_percent_GDP



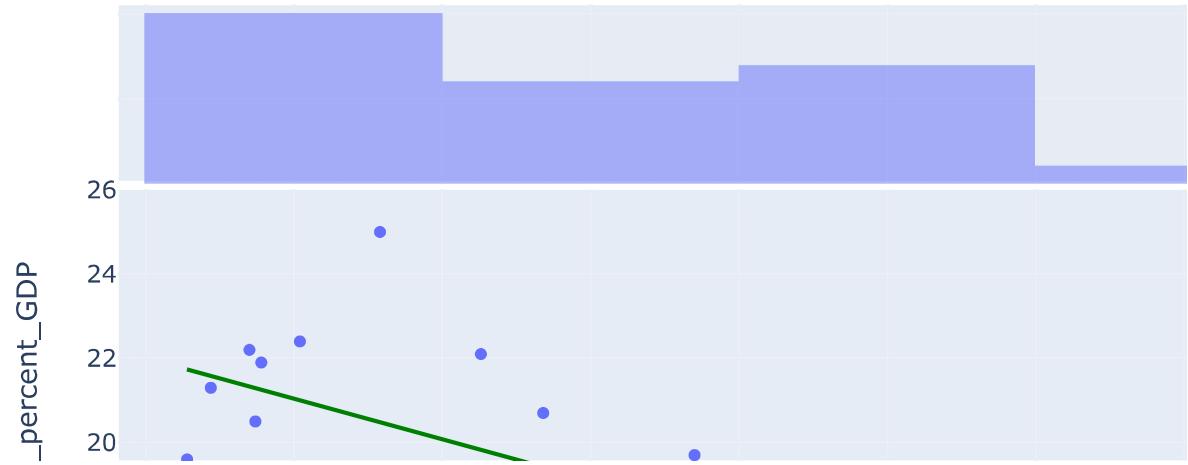
```
In [135]: corr_plot1(SS7_economy, 'Foreign_debt($)', 'Investment_percent_GDP', 'South Afric
```

Correlation between Foreign_debt(\$) and Investment_percent_GDP



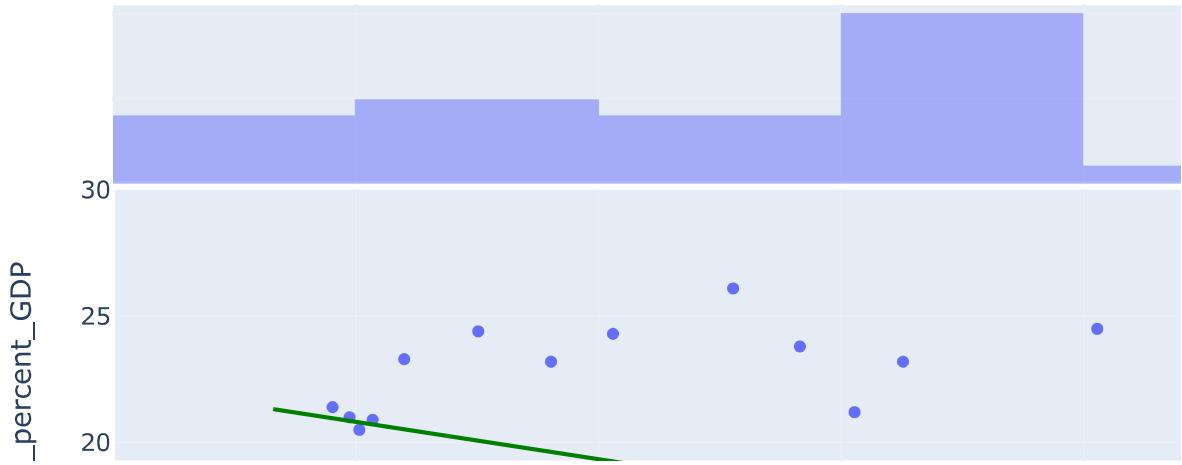
```
In [136]: corr_plot1(SS7_economy, 'Foreign_debt($)', 'Investment_percent_GDP', 'Kenya')
```

Correlation between Foreign_debt(\$) and Investment_percent_GDP



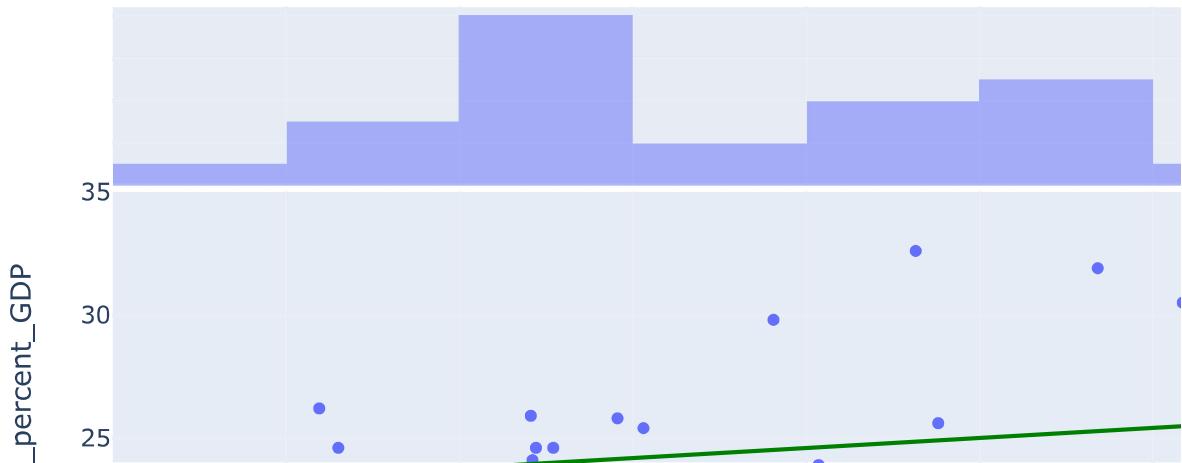
```
In [137]: corr_plot1(SS7_economy, 'Foreign_debt($)', 'Investment_percent_GDP', 'Rwanda')
```

Correlation between Foreign_debt(\$) and Investment_percent_GDP



```
In [138]: corr_plot1(SS7_economy, 'Foreign_debt($)', 'Investment_percent_GDP', 'Senegal')
```

Correlation between Foreign_debt(\$) and Investment_percent_GDP



for Kenya and Rwanda the higher the foreign_debt the lower the investment_percent_GDP

whereas Nigeria and Ghana the the higher the forien_debt the higher the investment_percent_GDP

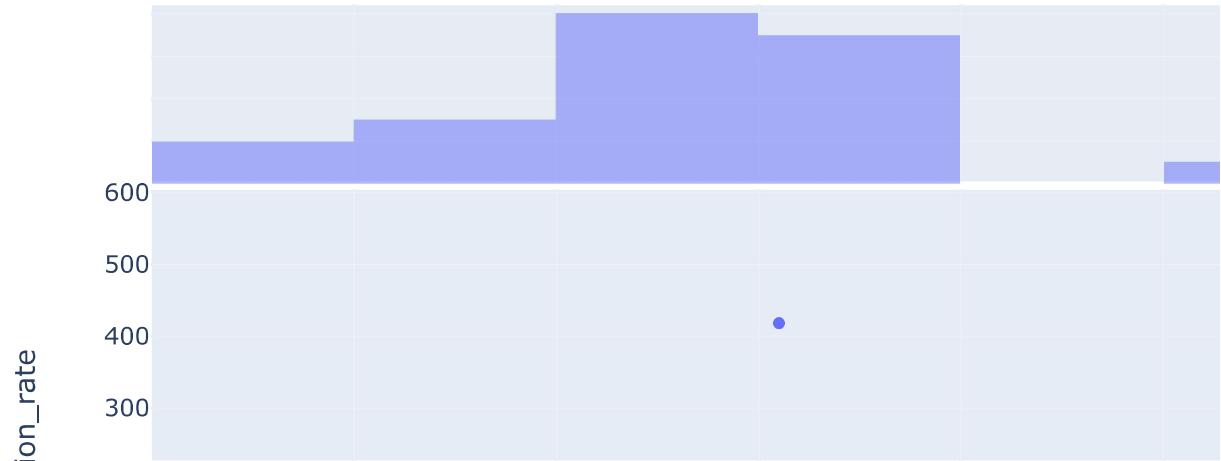
while South Africa and Senegal show no or very weak relationship between the Foreign_debt and Investment_percent_GDP

```
In [ ]:
```

Inflation vs Investment (G7)

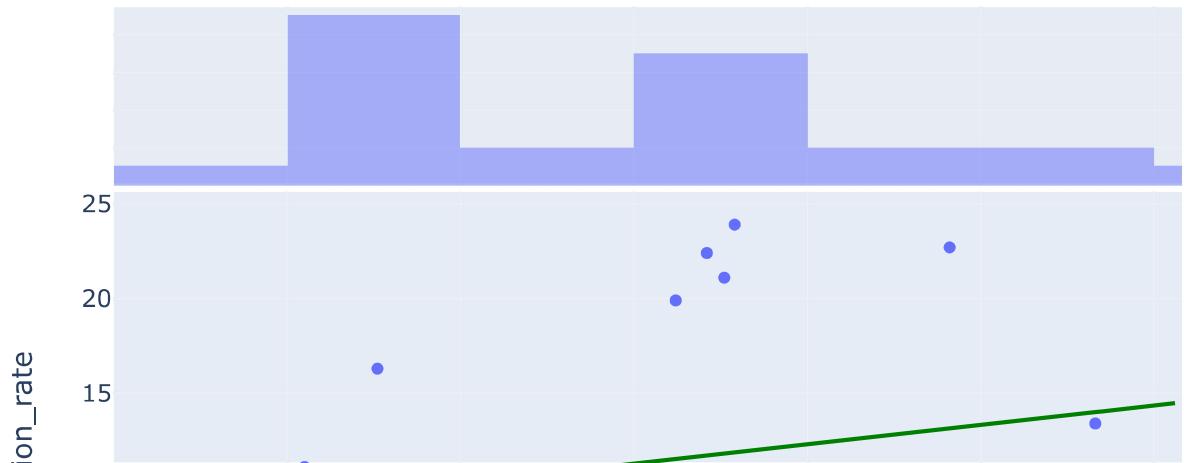
```
In [139]: corr_plot1(SS7_economy, 'Investment_percent_GDP', 'inflation_rate', 'Angola')
```

Correlation between Investment_percent_GDP and inflation_rate

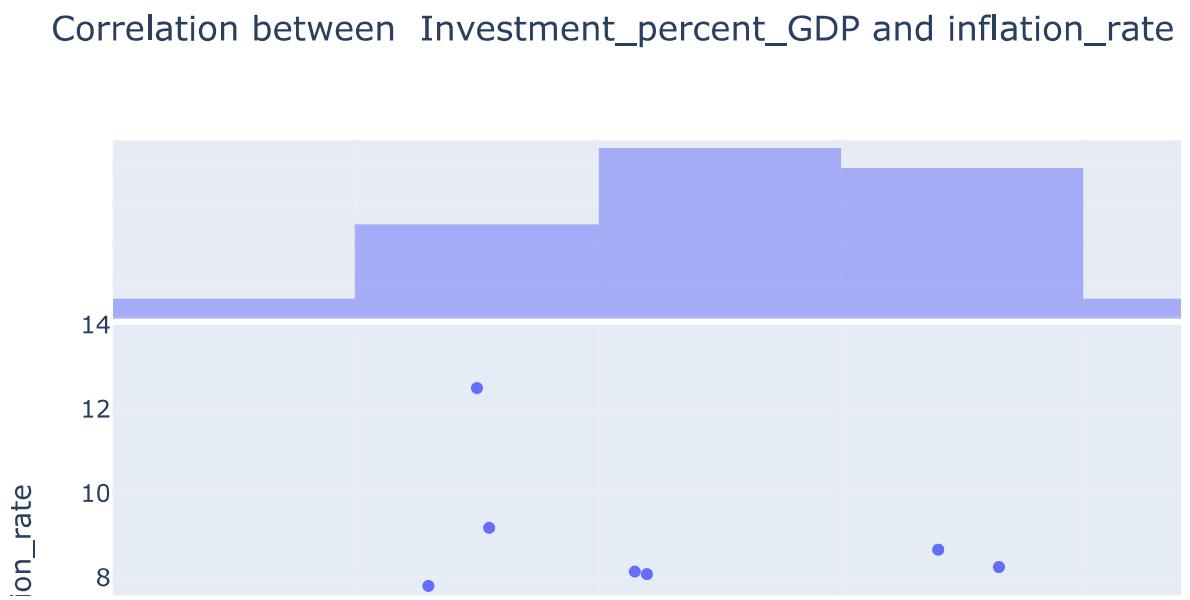


```
In [140]: corr_plot1(SS7_economy, 'Investment_percent_GDP', 'inflation_rate', 'Nigeria')
```

Correlation between Investment_percent_GDP and inflation_rate

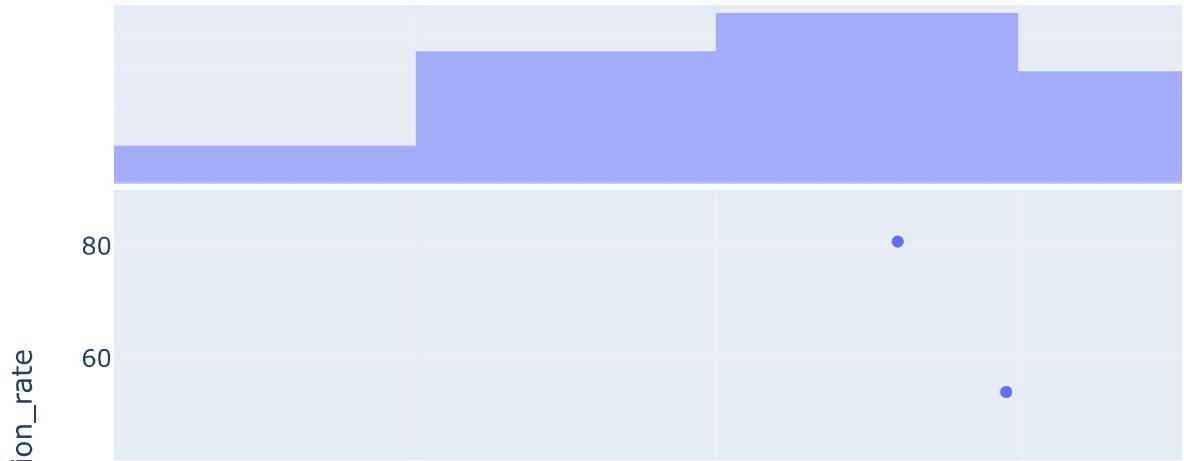


```
In [141]: corr_plot1(SS7_economy, 'Investment_percent_GDP', 'inflation_rate', 'South Africa')
```



```
In [142]: corr_plot1(SS7_economy, 'Investment_percent_GDP', 'inflation_rate', 'Ghana')
```

Correlation between Investment_percent_GDP and inflation_rate



```
In [143]: corr_plot1(SS7_economy, 'Investment_percent_GDP', 'inflation_rate', 'Senegal')
```

Correlation between Investment_percent_GDP and inflation_rate

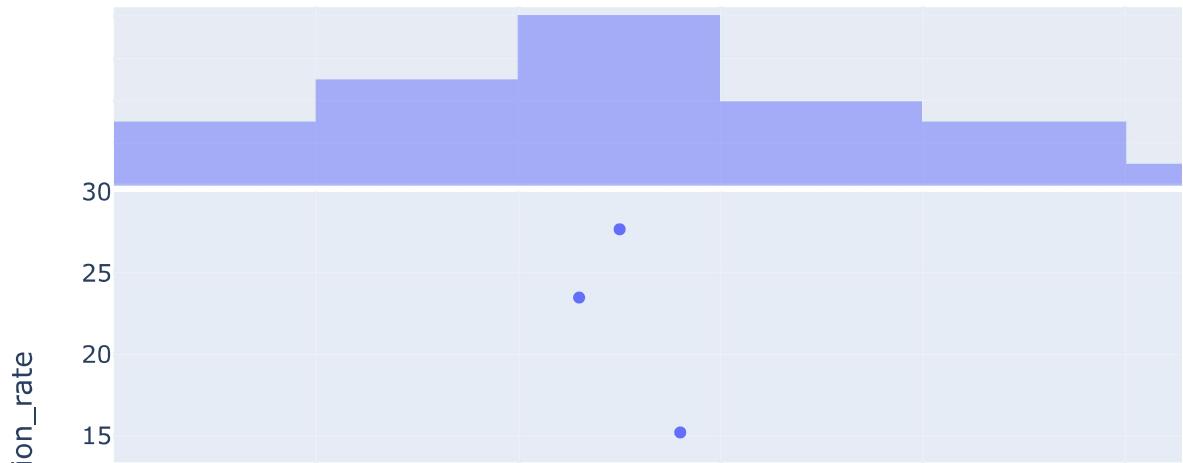


```
In [144]: corr_plot1(SS7_economy, 'Investment_percent_GDP', 'inflation_rate', 'Rwanda')
```



```
In [145]: corr_plot1(SS7_economy, 'Investment_percent_GDP', 'inflation_rate', 'Kenya')
```

Correlation between Investment_percent_GDP and inflation_rate



we can conclude that there is no significant relationship between inflation_rate and investment_percent_GDP amongst the top 7 sub saharan economy

```
In [ ]:
```

Trade_balance vs Investment

```
In [146]: corr_plot1(SS7_economy, 'Investment_percent_GDP', 'Trade_balance', 'Angola')
```

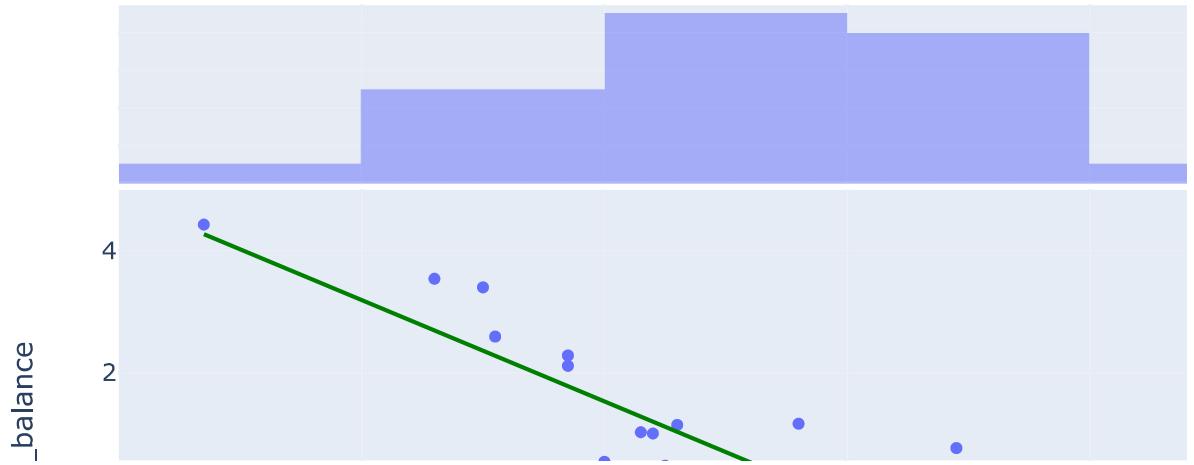


```
In [147]: corr_plot1(SS7_economy, 'Investment_percent_GDP', 'Trade_balance', 'Nigeria')
```



```
In [148]: corr_plot1(SS7_economy, 'Investment_percent_GDP', 'Trade_balance', 'South Africa')
```

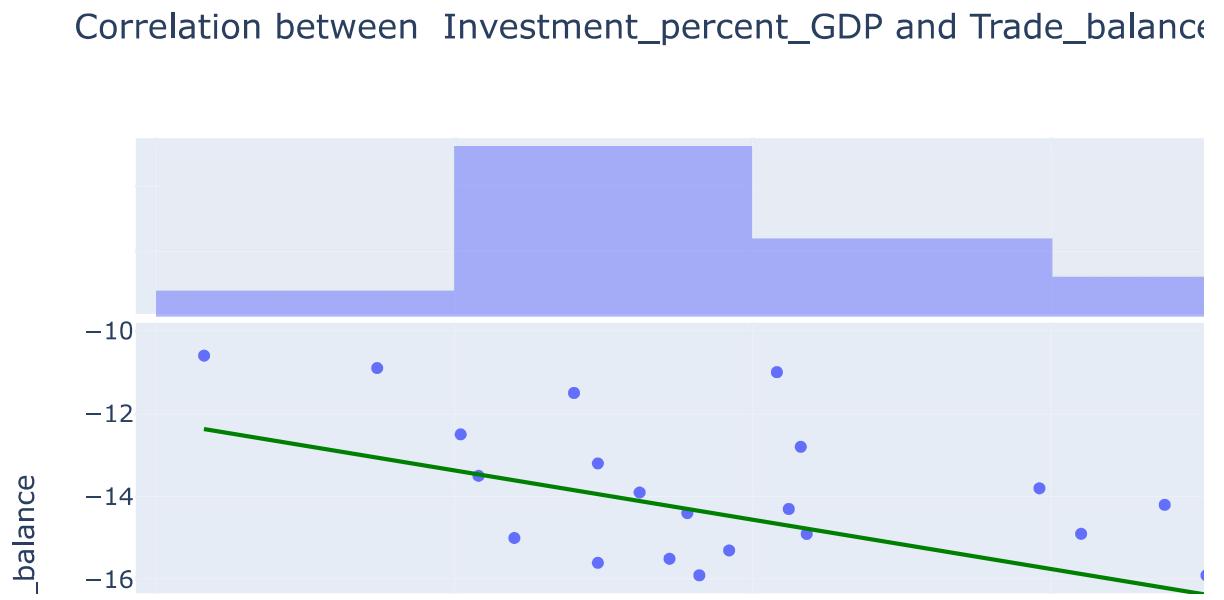
Correlation between Investment_percent_GDP and Trade_balance



```
In [149]: corr_plot1(SS7_economy, 'Investment_percent_GDP', 'Trade_balance', 'Ghana')
```



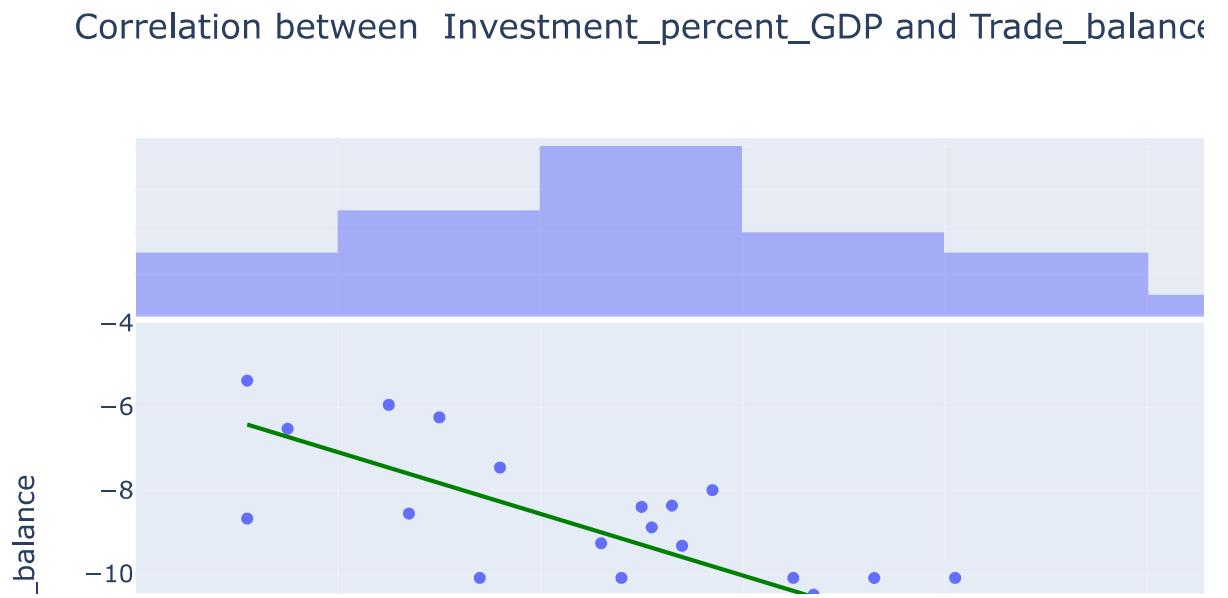
```
In [150]: corr_plot1(SS7_economy, 'Investment_percent_GDP', 'Trade_balance', 'Senegal')
```



```
In [151]: corr_plot1(SS7_economy, 'Investment_percent_GDP', 'Trade_balance', 'Rwanda')
```



```
In [152]: corr_plot1(SS7_economy, 'Investment_percent_GDP', 'Trade_balance', 'Kenya')
```



for Kenya, South Africa , Senegal and Rwanda the Higher the trade_balance the lower the investment_percent_GDP

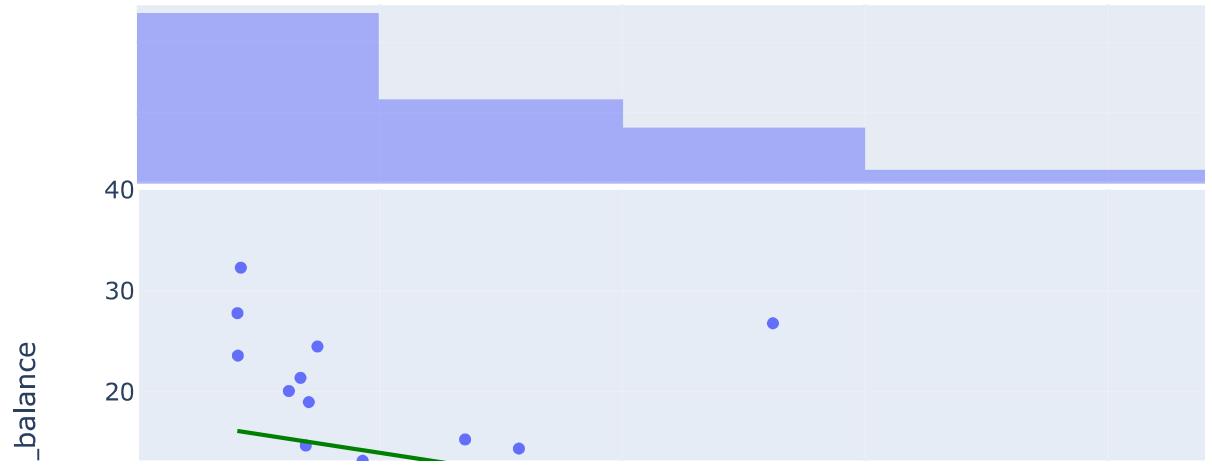
While Nigeria there's no significant relationship between trade_balance and Investment_percent_GDP

In []:

Foreign debt(\$) vs Trade balance

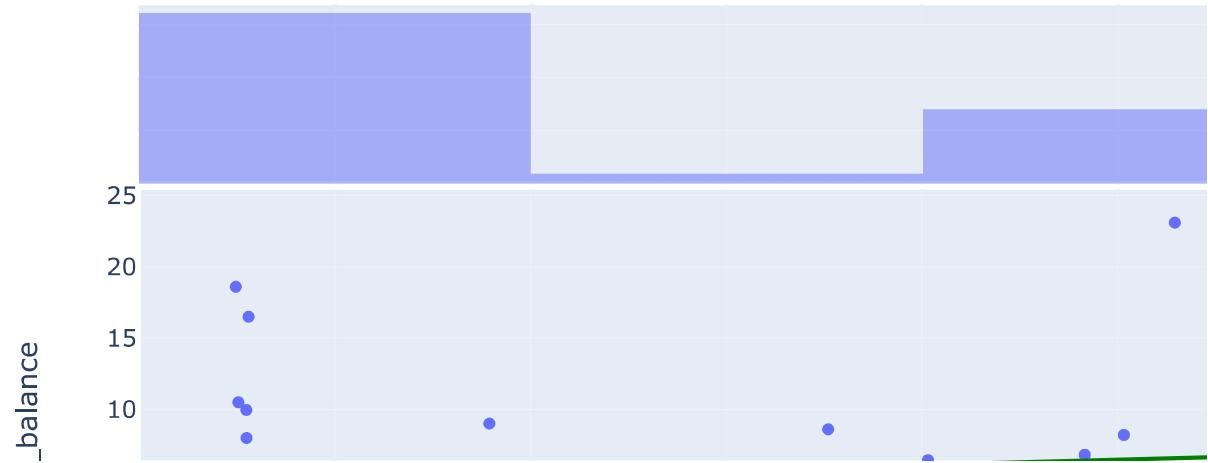
```
In [153]: corr_plot1(SS7_economy, 'Foreign_debt($)', 'Trade_balance', 'Angola')
```

Correlation between Foreign_debt(\$) and Trade_balance of Angol



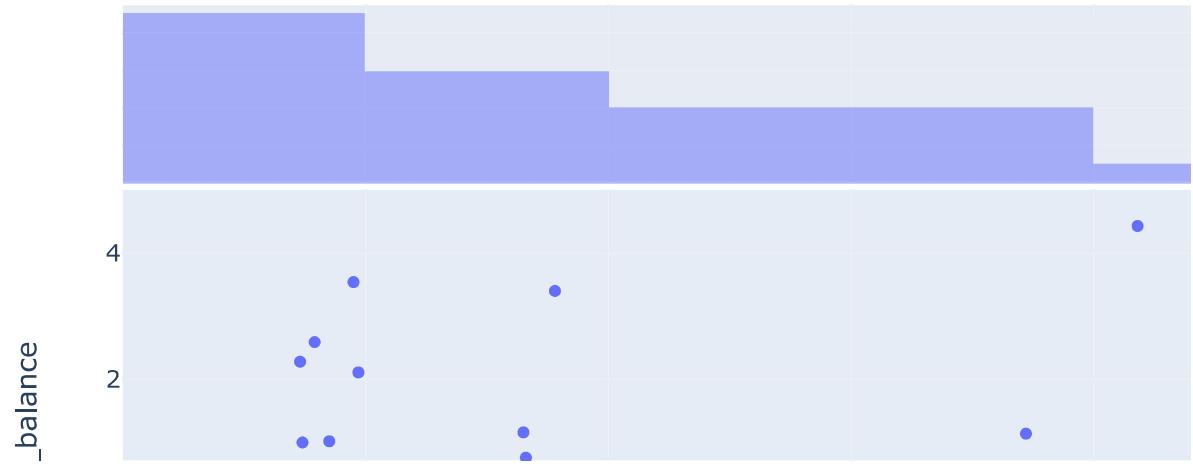
```
In [154]: corr_plot1(SS7_economy, 'Foreign_debt($)', 'Trade_balance', 'Nigeria')
```

Correlation between Foreign_debt(\$) and Trade_balance of Nigeria



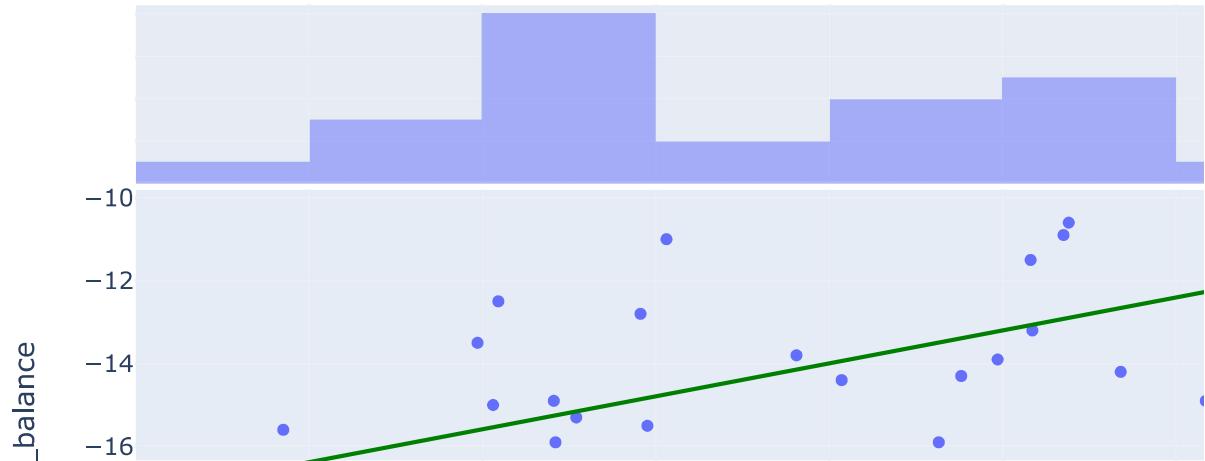
```
In [155]: corr_plot1(SS7_economy, 'Foreign_debt($)', 'Trade_balance', 'South Africa')
```

Correlation between Foreign_debt(\$) and Trade_balance of South Africa

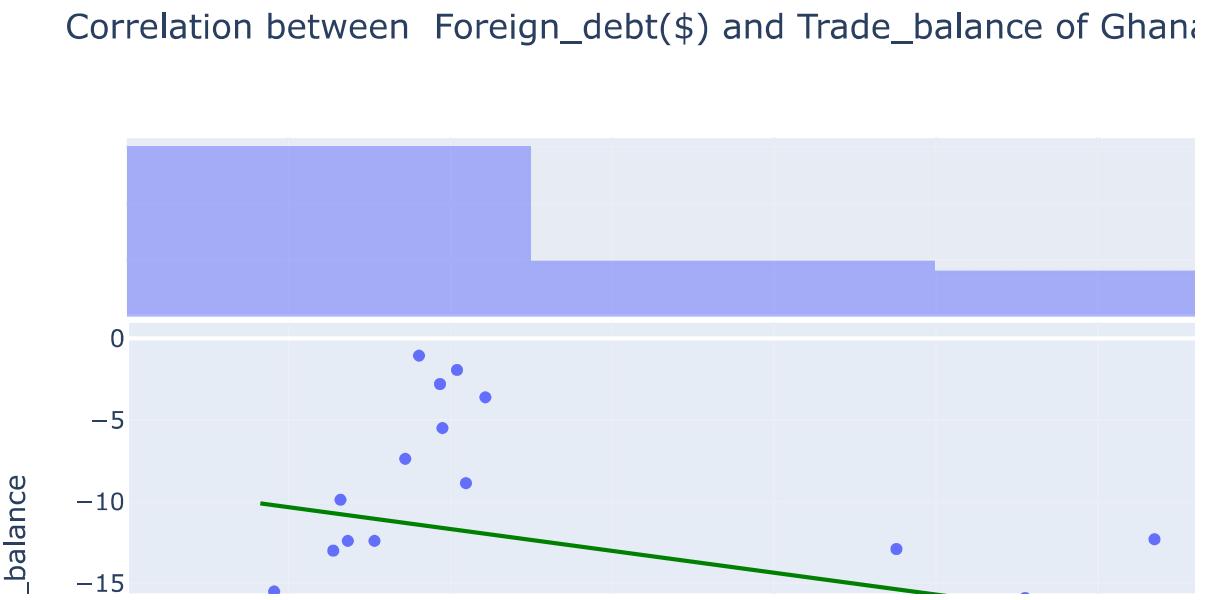


```
In [156]: corr_plot1(SS7_economy, 'Foreign_debt($)', 'Trade_balance', 'Senegal')
```

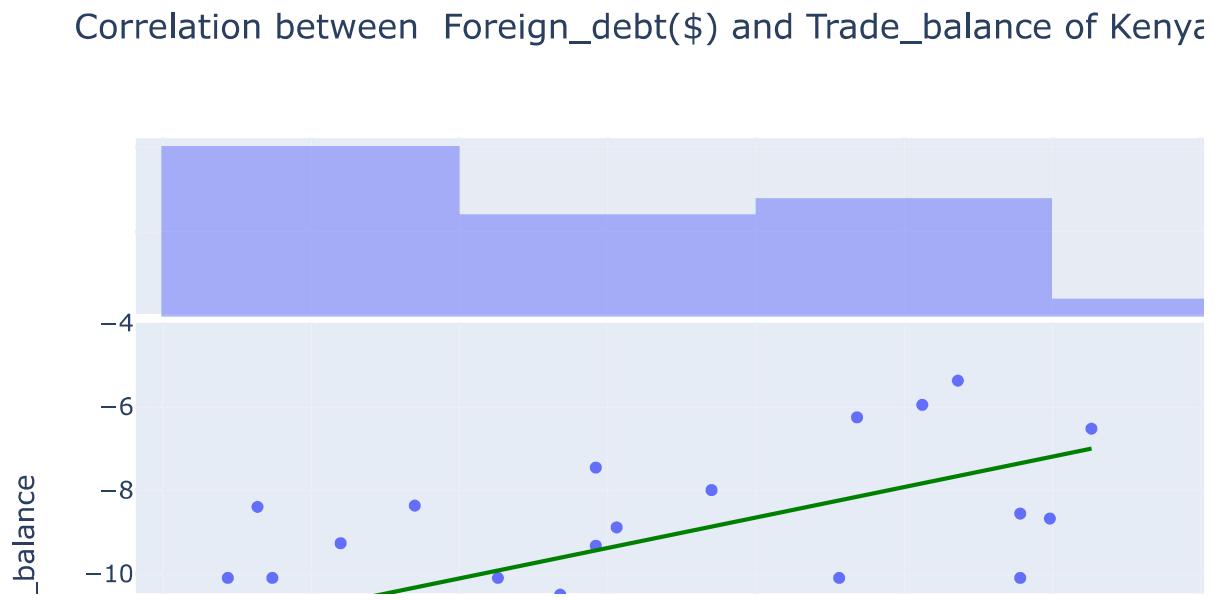
Correlation between Foreign_debt(\$) and Trade_balance of Senegal



```
In [157]: corr_plot1(SS7_economy, 'Foreign_debt($)', 'Trade_balance', 'Ghana')
```



```
In [158]: corr_plot1(SS7_economy, 'Foreign_debt($)', 'Trade_balance', 'Kenya')
```



for Kenya and Senegal the higher the foreign_debt the higher the trade_balance and vice versa

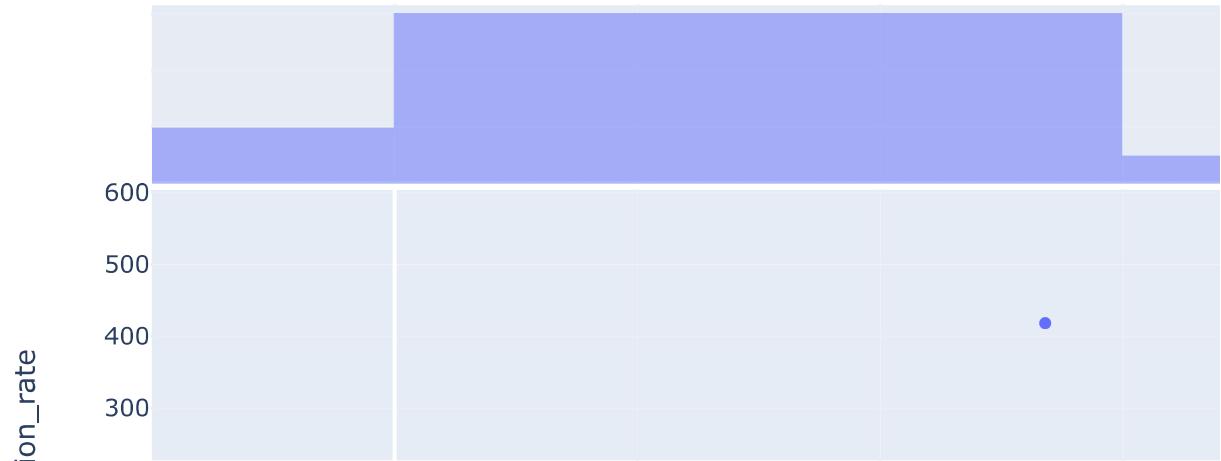
Whereas Ghana the higher the Foreign_debt the lower the trade_balance and vice versa

```
In [ ]:
```

Inflation vs Trade_balance

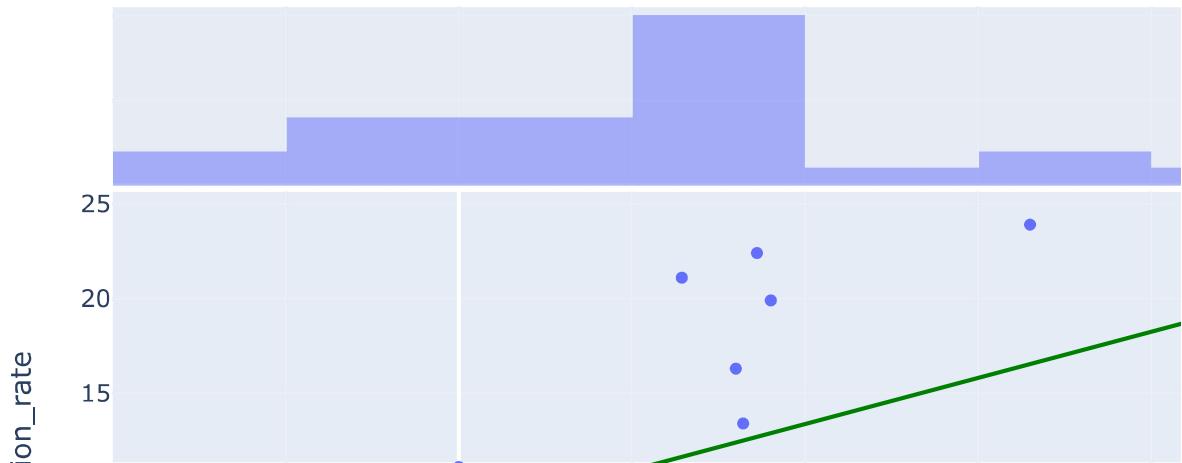
```
In [159]: corr_plot1(SS7_economy, 'Trade_balance', 'inflation_rate', 'Angola')
```

Correlation between Trade_balance and inflation_rate of Angola is:



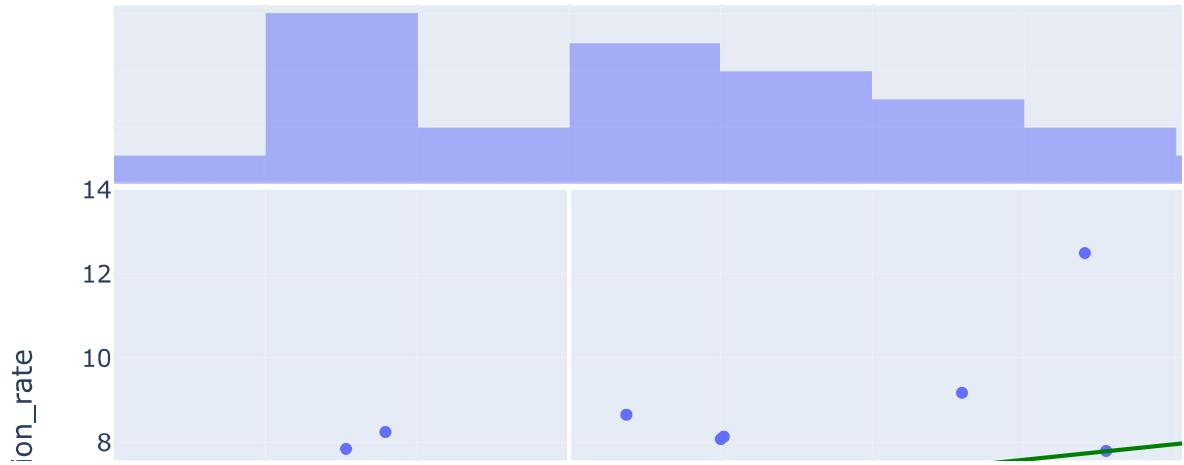
```
In [160]: corr_plot1(SS7_economy, 'Trade_balance', 'inflation_rate', 'Nigeria')
```

Correlation between Trade_balance and inflation_rate of Nigeria is:



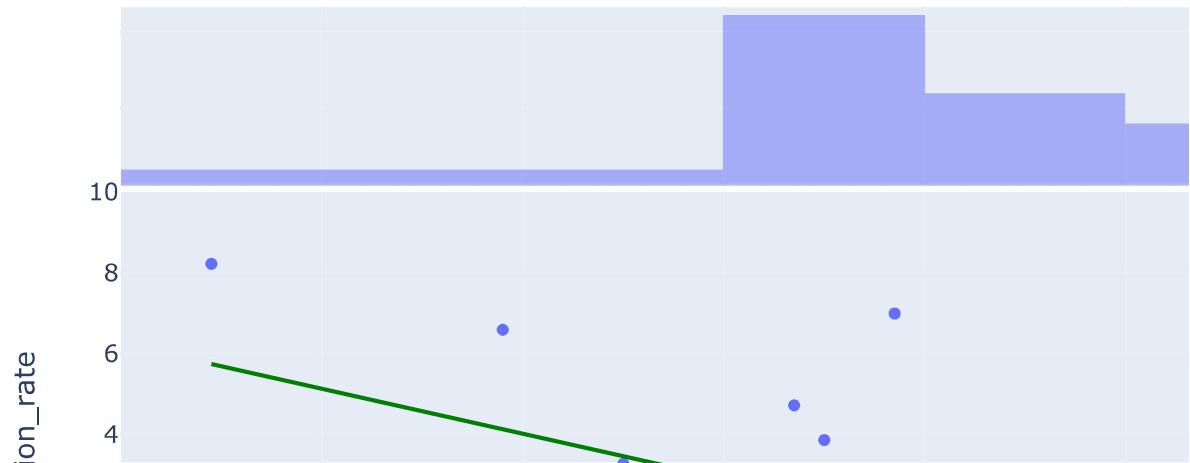
```
In [161]: corr_plot1(SS7_economy, 'Trade_balance', 'inflation_rate', 'South Africa')
```

Correlation between Trade_balance and inflation_rate of South Af



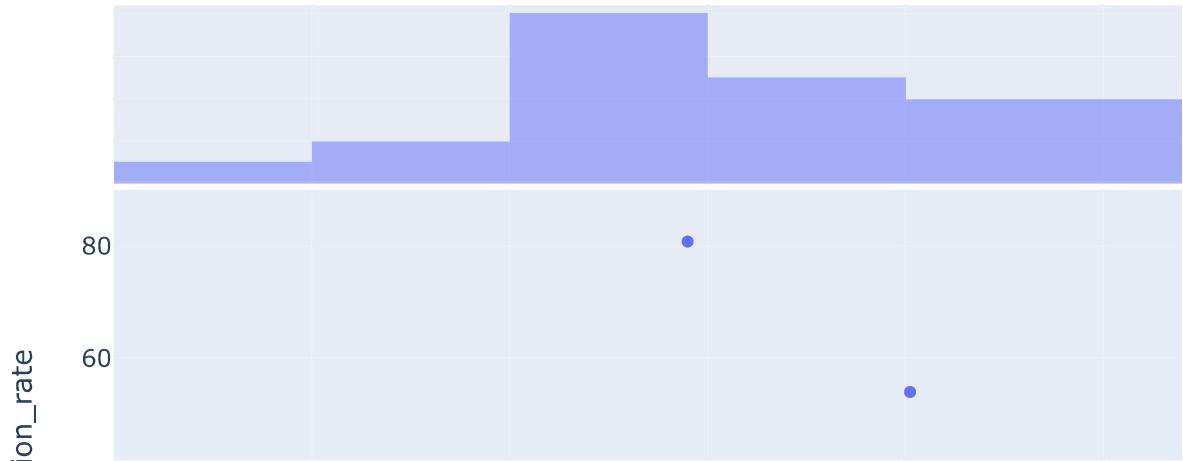
```
In [162]: corr_plot1(SS7_economy, 'Trade_balance', 'inflation_rate', 'Senegal')
```

Correlation between Trade_balance and inflation_rate of Senegal



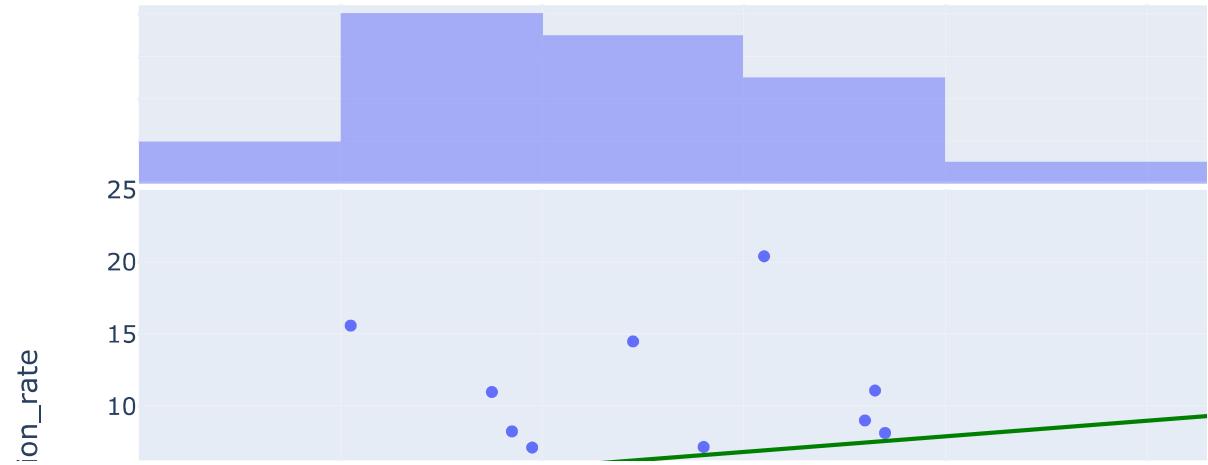
```
In [163]: corr_plot1(SS7_economy, 'Trade_balance', 'inflation_rate', 'Ghana')
```

Correlation between Trade_balance and inflation_rate of Ghana is



```
In [164]: corr_plot1(SS7_economy, 'Trade_balance', 'inflation_rate', 'Rwanda')
```

Correlation between Trade_balance and inflation_rate of Rwanda



```
In [165]: corr_plot1(SS7_economy, 'Trade_balance', 'inflation_rate', 'Kenya')
```

Correlation between Trade_balance and inflation_rate of Kenya is



For South Africa and Nigeria the relationship that exists between trade_balance and inflation_rate is negative amongst the top 7 sub saharan economies

Whereas Senegal it is the complete opposite

```
In [ ]:
```

Research answer 3:

3a. Yes it does vary as we can see from our analysis

[SS7_link](#)

3b. No it doesn't, some countries have a positive relationship while some have a negative relationship between trade_balance and Gross capital formation

[SS7_link](#)

[G7_link](#)

3c. The impact for G7 economies differs completely, while sub saharan economies there is no relationship

[G7_link](#)

[SS7_link](#)

3d. No it is not the same as we can see in our previous analysis

[SS7_link](#)

3e. No it is not the same

[SS7_link](#)

Conclusions

Results: Our data suggest that

1. For Countries like Nigeria and Ghana high foreign debt translate to high local investment (i.e investment_percent_GDP)

2. Some economic indicators are inter correlated with each other but differ from country to countries

3. Countries with similar economic structure like Nigeria and Angola (oil dependent economy) share some economic similarities and correlation

4. Foreign debt have no effect on Trade_balance on bigger economies in Sub Saharan Africa like Nigeria and South Africa

5. High inflation_rate seems to have a negative impact on investment_percent_GDP for a Country like Germany

6. Countries like the United States and Italy invested more when inflation was high

Limitations: There are couple of limitations in our data

1. Not having the Foreign_debt data for G7 countries have limited our analysis to just developing economies

2. In the absence of GDP growth data we have been unable to make more specific claims of how Foreign_debt, investment_percent_GDP , inflation_rate, and Trade_balance data effects GDP growth

In []:

In []: