Compare the birth rate in the Arab world to other continents

In [1]:
```
import pandas as pd
import numpy as np
import random
import matplotlib.pyplot as plt
```

World Development Indicators Dataset

In [2]:
```
data = pd.read_csv("./world-development-indicators/Indicators.csv")
```

In [3]:
```
data.isnull().any().any(), data.shape
```

Out[3]: (False, (5656459, 6))

Exploring Data

In [4]:
```
data.head(10)
```

Out[4]:

	CountryName	CountryCode	IndicatorName	IndicatorCode	Year	Value
0	Arab World	ARB	Adolescent fertility rate (births per 1,000 wo...	SP.ADO.TFRT	1960	1.33560943e+02
1	Arab World	ARB	Age dependency ratio (% of working-age populat...	SP.POP.DPND	1960	8.77976391e+01
2	Arab World	ARB	Age dependency ratio, old (% of working-age po...	SP.POP.DPND.OL	1960	6.63457repl00e+00
3	Arab World	ARB	Age dependency ratio, young (% of working-age ...	SP.POP.DPND.YG	1960	9.13233e+01
4	Arab World	ARB	Arms exports (SIPRI trend indicator values)	MS.MIL.XPRT.KD	1960	3.50000e+06
5	Arab World	ARB	Arms imports (SIPRI trend indicator values)	MS.MIL.MPRT.KD	1960	3.36000e+08
6	Arab World	ARB	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1960	4.76979e+01
7	Arab World	ARB	CO2 emissions (kt)	EN.ATM.CO2E.KT	1960	5.56639e+04
8	Arab World	ARB	CO2 emissions (metric tons per capita)	EN.ATM.CO2E.PC	1960	6.43982e-01
9	Arab World	ARB	CO2 emissions from gaseous fuel consumption (k...	EN.ATM.CO2E.GF	1960	5.04129e+00

In [5]:
```
countries = data["CountryName"].unique().tolist()
len(countries)
```

Out[5]: 247

In [6]:
```
countryCodes = data["CountryCode"].unique().tolist()
len(countryCodes)
```

Out[6]: 247

In [7]:
```
years = data["Year"].unique().tolist()
len(years)
```

Out[7]: 56

In [8]:
```
indicators = data["IndicatorName"].unique().tolist()
len(indicators)
```

Out[8]: 1344

Describe the data and find the unique in each columns

In [9]:
```
data.describe(include="all")
```

Out[9]:

	CountryName	CountryCode	IndicatorName	IndicatorCode	Year	Value
count	5656458	5656458	5656458	5656458	5.656458e+06	5.656458e+06
unique	247	247	1344	1344	NaN	NaN
top	Mexico	MEX	Population, total	SP.POP.TOTL	NaN	NaN
freq	37244	37244	13484	13484	NaN	NaN
mean	NaN	NaN	NaN	NaN	1.994484e+03	1.075501e+12
std	NaN	NaN	NaN	NaN	1.307589e+01	4.942409e+13
min	NaN	NaN	NaN	NaN	1.960000e+03	-8.824921e+15
25%	NaN	NaN	NaN	NaN	1.984000e+03	5.996242e+00
50%	NaN	NaN	NaN	NaN	1.997000e+03	8.357400e+01
75%	NaN	NaN	NaN	NaN	2.006000e+03	1.340722e+07
max	NaN	NaN	NaN	NaN	2.015000e+03	1.102397e+16

In [10]:
```
data[:20]["IndicatorName"]
```

Out[10]:
```
0     Adolescent fertility rate (births per 1,000 wo...
1     Age dependency ratio (% of working-age populat...
2     Age dependency ratio, old (% of working-age po...
3     Age dependency ratio, young (% of working-age ...
4          Arms exports (SIPRI trend indicator values)
5          Arms imports (SIPRI trend indicator values)
6                Birth rate, crude (per 1,000 people)
7                                  CO2 emissions (kt)
8              CO2 emissions (metric tons per capita)
9     CO2 emissions from gaseous fuel consumption (k...
10    CO2 emissions from liquid fuel consumption (kt)
11    CO2 emissions from liquid fuel consumption (k...
12    CO2 emissions from solid fuel consumption (k...
13                        Death rate, crude (per 1,000 people)
14                  Fertility rate, total (births per woman)
15                                 Fixed broadband subscriptions
16    Fixed telephone subscriptions (per 100 people)
17                                 Hospital beds (per 1,000 people)
18              International migrant stock (% of population)
19                   International migrant stock, total
Name: IndicatorName, dtype: object
```

In [11]:
```
print(min(years), "to ", max(years))
```

```
1960 to  2015
```

In [12]:
```
print(min(indicators), max(indicators))
```

```
2005 PPP conversion factor, GDP (LCU per international $) Youth literacy rate, population 15-24 years, male (%)
```

In [13]:
```
print(min(countries), max(countries))
```

```
Afghanistan Zimbabwe
```

Explore the Birth rate, crude (per 1,000) in ARB.

In [14]:
```
hist_indicator = 'Birth rate, crude \(per 1,000'
hist_country = 'ARB'

mask1 = data['IndicatorName'].str.contains(hist_indicator)
mask2 = data['CountryCode'].str.contains(hist_country)

stage = data[mask1 & mask2]
```

In [15]:
```
stage.head()
```

Out[15]:

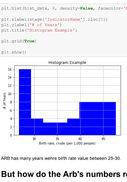
	CountryName	CountryCode	IndicatorName	IndicatorCode	Year	Value
0	Arab World	ARB	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1960	47.9978882
23200	Arab World	ARB	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1961	47.455318
46417	Arab World	ARB	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1962	47.198579
70256	Arab World	ARB	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1963	46.919370
106881	Arab World	ARB	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1964	46.612771
139928	Arab World	ARB	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1965	46.271651
168044	Arab World	ARB	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1966	45.891713
199301	Arab World	ARB	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1967	45.460863
232120	Arab World	ARB	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1968	45.000080
264467	Arab World	ARB	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1969	44.608645

Let's see how Birth rate have changed over time using MatplotLib

In [16]:
```
years = stage["Year"].values

bcr = stage["Value"].values

plt.bar(years,bcr)
plt.show()
```

To make more honest, start the y axis at 0

In [17]:
```
plt.plot(stage["Year"].values, stage["Value"].values)

plt.xlabel("Year")
plt.ylabel(stage["IndicatorName"].iloc[0])

plt.title(" Birth rate, crude (per 1,000)")

plt.axis([1959, 2013,0,60])
plt.show()
```



Find the mean of Birth rate, crude (per 1,000) from 1960 to 2013

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

Out[18]:

	Year	Value
count	54.000000	54.000000
mean	1986.500000	36.692245
std	15.732133	7.482920
min	1960.000000	20.704356
25%	1973.250000	28.377561
50%	1986.500000	38.029148
75%	1999.750000	43.812388
max	2013.000000	47.697888

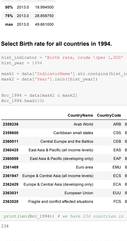
In [19]:
```
hist_data = stage["Value"].values
```

In [20]:
```
print(len(hist_data))
```

```
54
```

Using Histograms to explore the distribution of values

In [21]:
```
plt.hist(hist_data, 7, density=False, facecolor='blue')

plt.xlabel(stage["IndicatorName"].iloc[0])
plt.ylabel('# of Years')
plt.title('Histogram Example')

plt.grid(True)

plt.show()
```



ARB has many years wehre birth rate value between 25-30.

But how do the Arb's numbers relate to those of other countries?

Select Birth rate for all countries in 2013.

In [22]:
```
hist_indicator = 'Birth rate, crude \(per 1,000'
hist_year = 2013

mask1 = data['IndicatorName'].str.contains(hist_indicator)
mask2 = data['Year'].isin([hist_year])


Brc_2013 = data[mask1 & mask2]
Brc_2013.head(10)
```

Out[22]:

	CountryName	CountryCode	IndicatorName	IndicatorCode	Year	Value
5377473	Arab World	ARB	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	2013	26.704266
5377845	Caribbean small states	CSS	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	2013	15.495534
5378387	Central Europe and the Baltics	CEB	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	2013	9.605113
5378916	East Asia & Pacific (all income levels)	EAS	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	2013	13.737584
5379462	East Asia & Pacific (developing only)	EAP	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	2013	14.316272
5380198	Euro area	EMU	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	2013	9.71992
5380724	Europe & Central Asia (all income levels)	ECS	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	2013	12.351336
5381340	Europe & Central Asia (developing only)	ECA	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	2013	16.118729
5382133	European Union	EUU	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	2013	10.014735
5382868	Fragile and conflict affected situations	FCS	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	2013	33.533332

For how many countries do we have Birth rate data in 2013

In [23]:
```
print(len(Brc_2013))
```

```
234
```

The mean value for all countries birth rate in 2013

In [24]:
```
Brc_2013.mean()
```

Out[24]:
```
Year      2013.000000
Value       21.188047
dtype: float64
```

In [25]:
```
Brc_2013.describe()
```

Out[25]:

	Year	Value
count	234.0	234.000000
mean	2013.0	21.188047
std	0.0	10.312817
min	2013.0	7.900000
25%	2013.0	12.226500
50%	2013.0	18.994000
75%	2013.0	28.808750
max	2013.0	49.661000

Select Birth rate for all countries in 1994.

In [26]:
```
hist_indicator = 'Birth rate, crude \(per 1,000'
hist_year = 1994

mask1 = data['IndicatorName'].str.contains(hist_indicator)
mask2 = data['Year'].isin([hist_year])


Brc_1994 = data[mask1 & mask2]
Brc_1994.head(10)
```

Out[26]:

	CountryName	CountryCode	IndicatorName	IndicatorCode	Year	Value
2358239	Arab World	ARB	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1994	31.921563
2358600	Caribbean small states	CSS	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1994	23.235537
2358911	Central Europe and the Baltics	CEB	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1994	11.328565
2360428	East Asia & Pacific (all income levels)	EAS	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1994	18.606421
2360859	East Asia & Pacific (developing only)	EAP	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1994	19.738568
2361469	Euro area	EMU	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1994	10.563035
2361947	Europe & Central Asia (all income levels)	ECS	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1994	12.825229
2362429	Europe & Central Asia (developing only)	ECA	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1994	18.201256
2363031	European Union	EUU	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1994	11.082733
2363520	Fragile and conflict affected situations	FCS	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1994	39.363358

In [27]:
```
print(len(Brc_1994)) # we have 234 countries in 1994.
```

```
234
```

The mean value for all countries birth rate in 1994

In [28]:
```
Brc_1994.describe()
```

Out[28]:

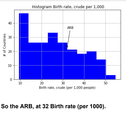
	Year	Value
count	234.0	234.000000
mean	1994.0	26.621010
std	0.0	11.904197
min	1994.0	8.400000
25%	1994.0	15.962500
50%	1994.0	24.974500
75%	1994.0	35.724000
max	1994.0	54.818000

In [29]:
```
Brc_1994.mean()
```

Out[29]:
```
Year      1994.00000
Value       26.62101
dtype: float64
```

let's plot a histogram of the birth rate per 1000 by country

In [30]:
```
fig, ax = plt.subplots()

ax.annotate("ARB",
            xy=(33, 22), xycoords='data',
            xytext=(-150, -140), textcoords='data',
            arrowprops=dict(facecolor="r",
                            connectionstyle="arc3"),
            )

plt.hist(Brc_1994["Value"], 10, density=False, facecolor='blue')

plt.xlabel(stage["IndicatorName"].iloc[0])
plt.ylabel('# of Countries')
plt.title('Histogram Birth rate, crude per 1,000')

#plt.axis([25, 32, 0, 14])
plt.grid(True)

plt.show()
```



So the ARB, at 32 Birth rate (per 1000).

In [ ]:
```

```