

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, (5656458, 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...	SPADO.TFR.T	1960	1.335809e+02
1	Arab World	ARB	Age dependency ratio (% of working-age populat...	SPPOCPDPND	1960	8.776760e+01
2	Arab World	ARB	Age dependency ratio, old (% of working-age po...	SPPOCPDPND.OL	1960	8.634579e+00
3	Arab World	ARB	Age dependency ratio, young (% of working-age ...	SPPOCPDPND.YG	1960	8.102333e+01
4	Arab World	ARB	Arms exports (SIPRI trend indicator values)	MS.MIL.XPRTKD	1960	3.000000e+06
5	Arab World	ARB	Arms imports (SIPRI trend indicator values)	MS.MIL.MPRTKD	1960	5.380000e+06
6	Arab World	ARB	Birth rate, crude (per 1,000 people)	SPDYN.CBRTIN	1960	4.769789e+01
7	Arab World	ARB	CO2 emissions (kt)	EN.ATM.CO2E.KT	1960	5.956399e+04
8	Arab World	ARB	CO2 emissions (metric tons per capita)	EN.ATM.CO2E.PC	1960	6.439635e-01
9	Arab World	ARB	CO2 emissions from gaseous fuel consumption (%...	EN.ATM.CO2E.GF.ZS	1960	5.041292e+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	SPPOPTOTL	NaN	NaN
freq	37244	37244	13484	13484	NaN	NaN
mean	NaN	NaN	NaN	NaN	1.994464e+03	1.070501e+12
std	NaN	NaN	NaN	NaN	1.387895e+01	4.842469e+13
min	NaN	NaN	NaN	NaN	1.960000e+03	-9.824821e+15
25%	NaN	NaN	NaN	NaN	1.984000e+03	5.566242e+00
50%	NaN	NaN	NaN	NaN	1.997000e+03	6.357450e+01
75%	NaN	NaN	NaN	NaN	2.006000e+03	1.346722e+07
max	NaN	NaN	NaN	NaN	2.015000e+03	1.103367e+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 (...
10 CO2 emissions from liquid fuel consumption (% ...
11 CO2 emissions from liquid fuel consumption (kt)
12 CO2 emissions from solid fuel consumption (% o...
13 Death rate, crude (per 1,000 people)
14 Fertility rate, total (births per woman)
15 Fixed telephone 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(10)

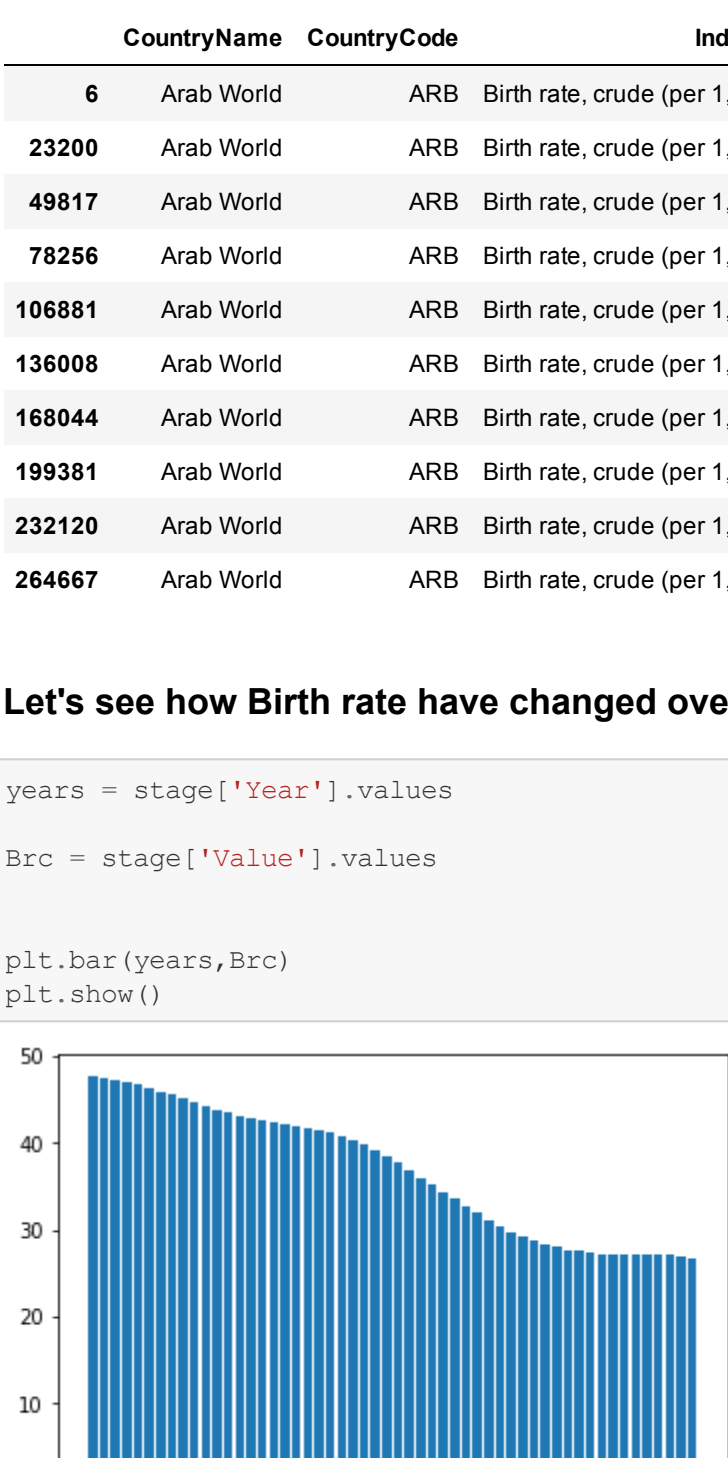
Out[15]:
```

	CountryName	CountryCode	IndicatorName	IndicatorCode	Year	Value
6	Arab World	ARB	Birth rate, crude (per 1,000 people)	SPDYN.CBRTIN	1960	47.697888
23200	Arab World	ARB	Birth rate, crude (per 1,000 people)	SPDYN.CBRTIN	1961	47.455318
49817	Arab World	ARB	Birth rate, crude (per 1,000 people)	SPDYN.CBRTIN	1962	47.198579
78256	Arab World	ARB	Birth rate, crude (per 1,000 people)	SPDYN.CBRTIN	1963	46.919370
106881	Arab World	ARB	Birth rate, crude (per 1,000 people)	SPDYN.CBRTIN	1964	46.612771
136008	Arab World	ARB	Birth rate, crude (per 1,000 people)	SPDYN.CBRTIN	1965	46.271631
168044	Arab World	ARB	Birth rate, crude (per 1,000 people)	SPDYN.CBRTIN	1966	45.891713
199381	Arab World	ARB	Birth rate, crude (per 1,000 people)	SPDYN.CBRTIN	1967	45.480863
232120	Arab World	ARB	Birth rate, crude (per 1,000 people)	SPDYN.CBRTIN	1968	45.050089
264667	Arab World	ARB	Birth rate, crude (per 1,000 people)	SPDYN.CBRTIN	1969	44.608845

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

```
In [16]: years = stage['Year'].values
Brc = stage['Value'].values

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



To make more honest, start the y axis at 0

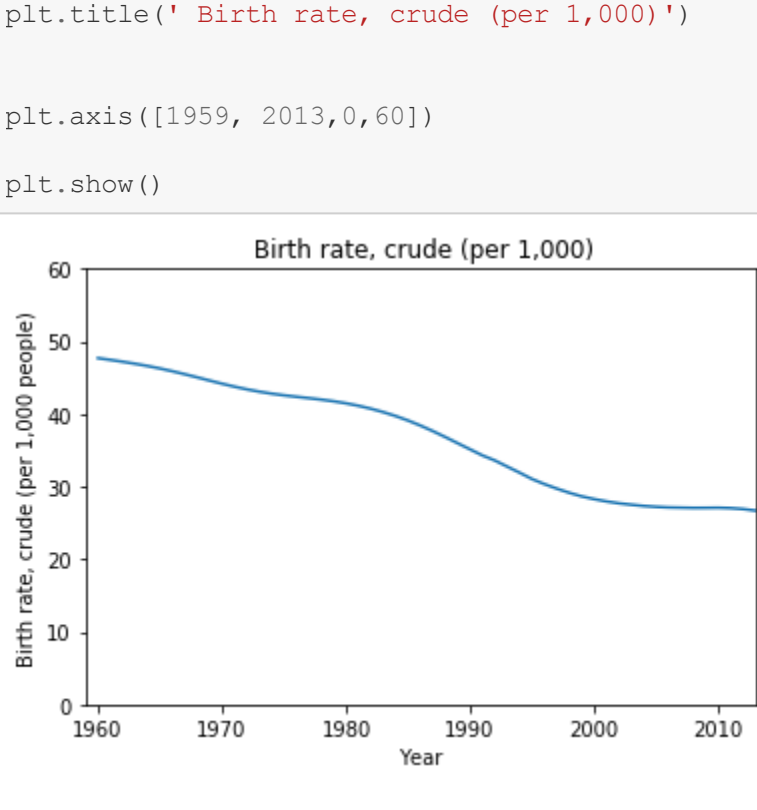
```
In [17]: plt.plot(stage['Year'].values, stage['Value'].values)

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	36.683245
std	15.732133	7.452920
min	1960.000000	26.704356
25%	1973.250000	28.377051
50%	1986.500000	38.030146
75%	1999.750000	43.012388
max	2013.000000	47.697888

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

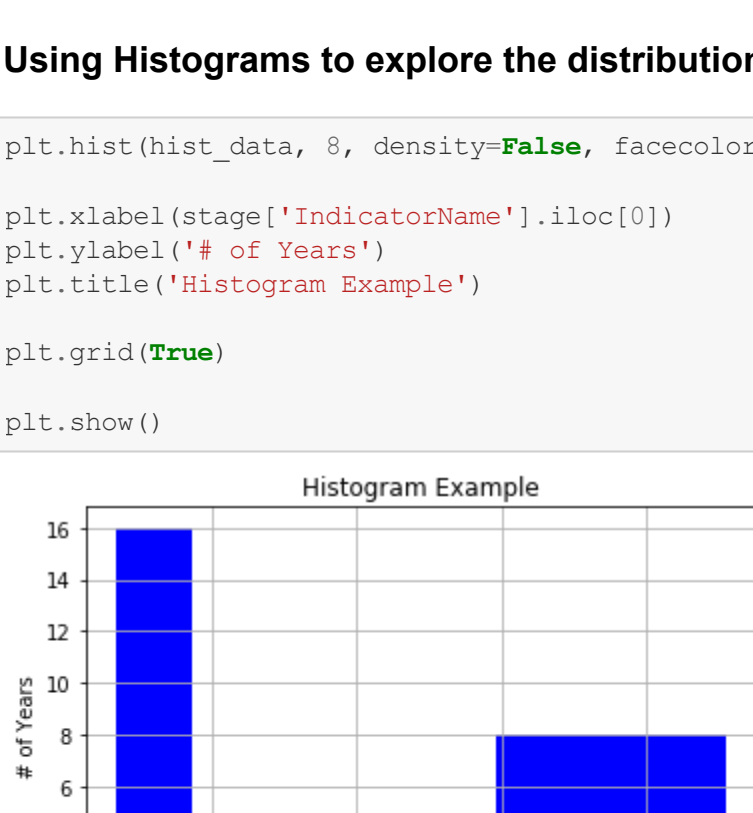
In [20]: print(len(hist_data))

54
```

Using Histograms to explore the distribution of values

```
In [21]: plt.hist(hist_data, 8, 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)	SPDYN.CBRTIN	2013	26.704356
5377945	Caribbean small states	CSS	Birth rate, crude (per 1,000 people)	SPDYN.CBRTIN	2013	15.469034
5378387	Central Europe and the Baltics	CEB	Birth rate, crude (per 1,000 people)	SPDYN.CBRTIN	2013	9.505113
5378916	East Asia & Pacific (all income levels)	EAS	Birth rate, crude (per 1,000 people)	SPDYN.CBRTIN	2013	13.737894
5379462	East Asia & Pacific (developing only)	EAP	Birth rate, crude (per 1,000 people)	SPDYN.CBRTIN	2013	14.318272
5380195	Euro area	EMU	Birth rate, crude (per 1,000 people)	SPDYN.CBRTIN	2013	9.710992
5380724	Europe & Central Asia (all income levels)	ECS	Birth rate, crude (per 1,000 people)	SPDYN.CBRTIN	2013	12.351338
5381340	Europe & Central Asia (developing only)	ECA	Birth rate, crude (per 1,000 people)	SPDYN.CBRTIN	2013	16.116729
5382135	European Union	EUU	Birth rate, crude (per 1,000 people)	SPDYN.CBRTIN	2013	10.014735
5382668	Fragile and conflict affected situations	FCS	Birth rate, crude (per 1,000 people)	SPDYN.CBRTIN	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	10.313917
min	2013.0	7.900000
25%	2013.0	12.229000
50%	2013.0	18.994500
75%	2013.0	28.858750
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
2359238	Arab World	ARB	Birth rate, crude (per 1,000 people)	SPDYN.CBRTIN	1994	31.921923
2359600	Caribbean small states	CSS	Birth rate, crude (per 1,000 people)	SPDYN.CBRTIN	1994	23.215537
2360011	Central Europe and the Baltics	CEB	Birth rate, crude (per 1,000 people)	SPDYN.CBRTIN	1994	11.326650
2360425	East Asia & Pacific (all income levels)	EAS	Birth rate, crude (per 1,000 people)	SPDYN.CBRTIN	1994	16.900421
2360859	East Asia & Pacific (developing only)	EAP	Birth rate, crude (per 1,000 people)	SPDYN.CBRTIN	1994	19.758868
2361469	Euro area	EMU	Birth rate, crude (per 1,000 people)	SPDYN.CBRTIN	1994	10.563035
2362724	Europe & Central Asia (all income levels)	ECS	Birth rate, crude (per 1,000 people)	SPDYN.CBRTIN	1994	15.351338
2362429	Europe & Central Asia (developing only)	ECA	Birth rate, crude (per 1,000 people)	SPDYN.CBRTIN	1994	18.201250
2363031	European Union	EUU	Birth rate, crude (per 1,000 people)	SPDYN.CBRTIN	1994	11.082733
2363520	Fragile and conflict affected situations	FCS	Birth rate, crude (per 1,000 people)	SPDYN.CBRTIN	1994	39.383358

```
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	11.904107
min	1994.0	9.400000
25%	1994.0	15.962900
50%	1994.0	24.974500
75%	1994.0	35.724500
max	1994.0	54.816000

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

Out[29]: Year 1994.000000
Value 26.621010
dtype: float64
```

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

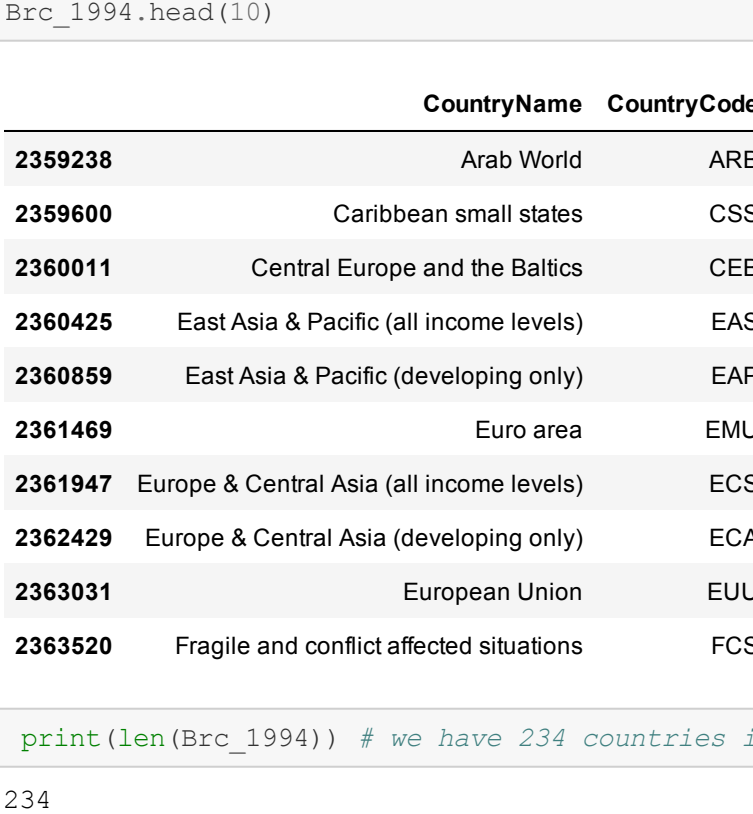
```
In [30]: Fig, ax = plt.subplots()

ax.annotate("ARB",
           xy=(32, 22), xycoords='data',
           xytext=(32, 36), textcoords='data',
           arrowprops=dict(arrowstyle="->",
                           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([10, 22, 0, 14])
plt.grid(True)
plt.show()
```



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

No References.

```
In [ ]:
```