Mini Project Week 6

September 3, 2018

World Development Indicators

Exploring Indicator of "Death Rate", "Hospital Beds", "Maternal Mortality" and "Government expenditure on education" for Argentina and Brazil

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

In [2]:
```
data = pd.read_csv('./Indicators-Copy1.csv')
data.shape
```

Out[2]:
```
(5656458, 6)
```

In [3]:
```
#Chosing the indicators
indicators = data['IndicatorName'].unique().tolist()
#indicators
```

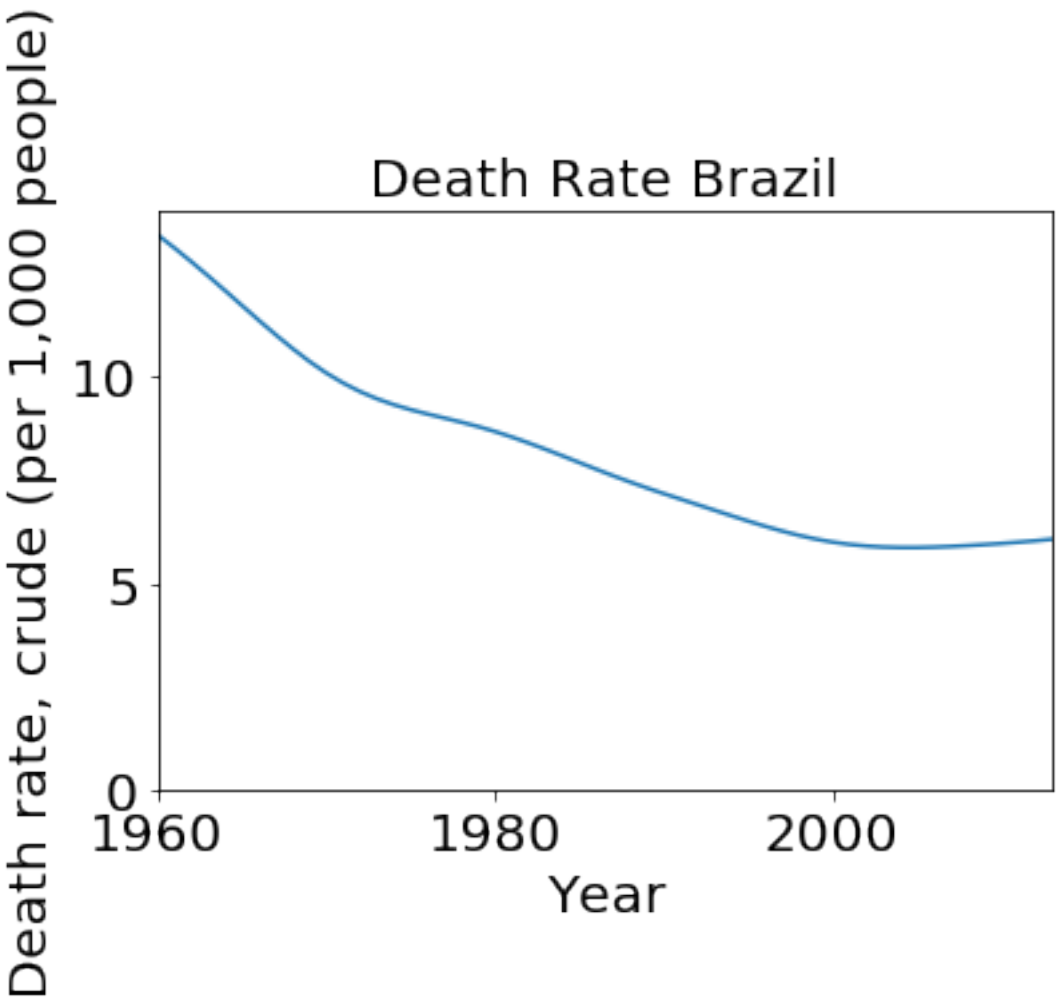
In [4]:
```
# select Death Rate, Hospital Beds, Maternal Mortality and Government expenditure on e

brazil = data['CountryCode'].str.contains('BRA')
indicatorDeathRate = data['IndicatorName'].str.contains("Death rate, crude")
indicatorHospitalBeds = data['IndicatorName'].str.contains("Hospital beds")
indicatorMaternalMortality = data['IndicatorName'].str.contains("Maternal mortality ra
indicatorGovExpEdu = data['IndicatorName'].str.contains("Government expenditure on edu

# indicators matching the BRA for country code and the choosed indicators over time.
dataBrDeath = data[brazil & indicatorDeathRate]
dataBrHosp = data[brazil & indicatorHospitalBeds]
dataBrMaternalMort = data[brazil & indicatorMaternalMortality]
dataBrGovExpEdu = data[brazil & indicatorGovExpEdu]
```

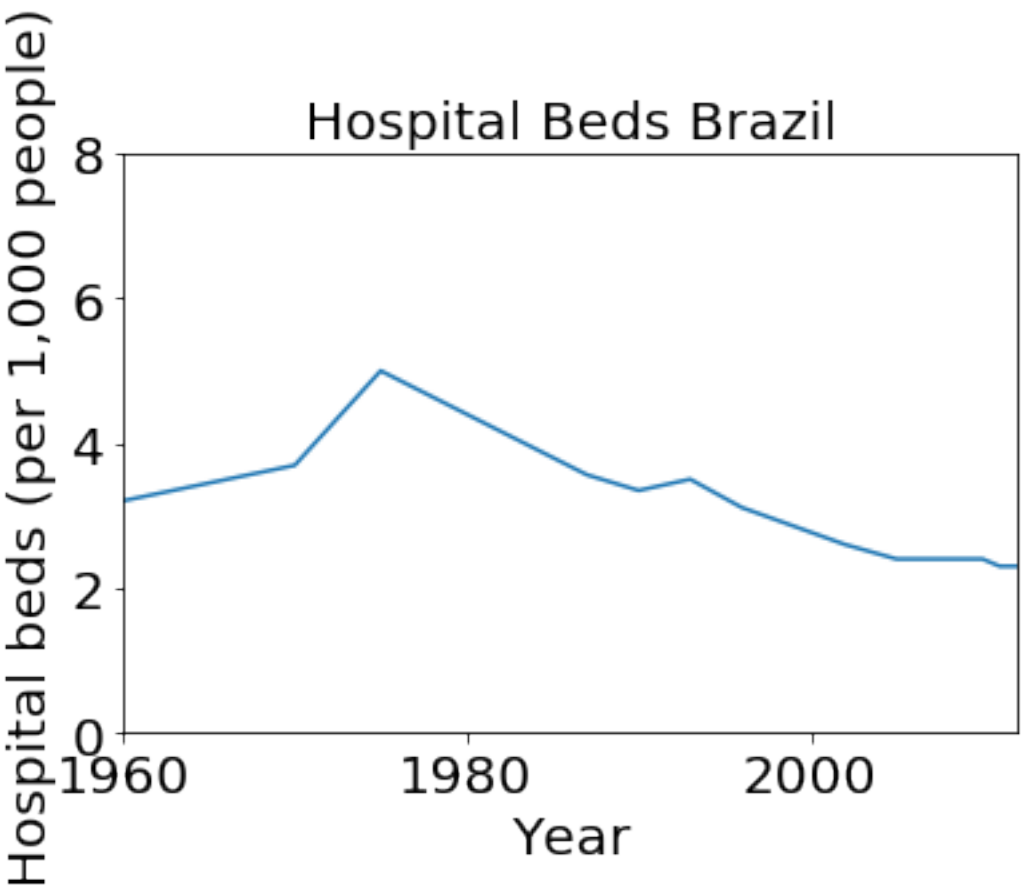
In [5]:
```
#see the size of the indicators
dataBrDeath.shape, dataBrHosp.shape, dataBrMaternalMort.shape, dataBrGovExpEdu.shape
```

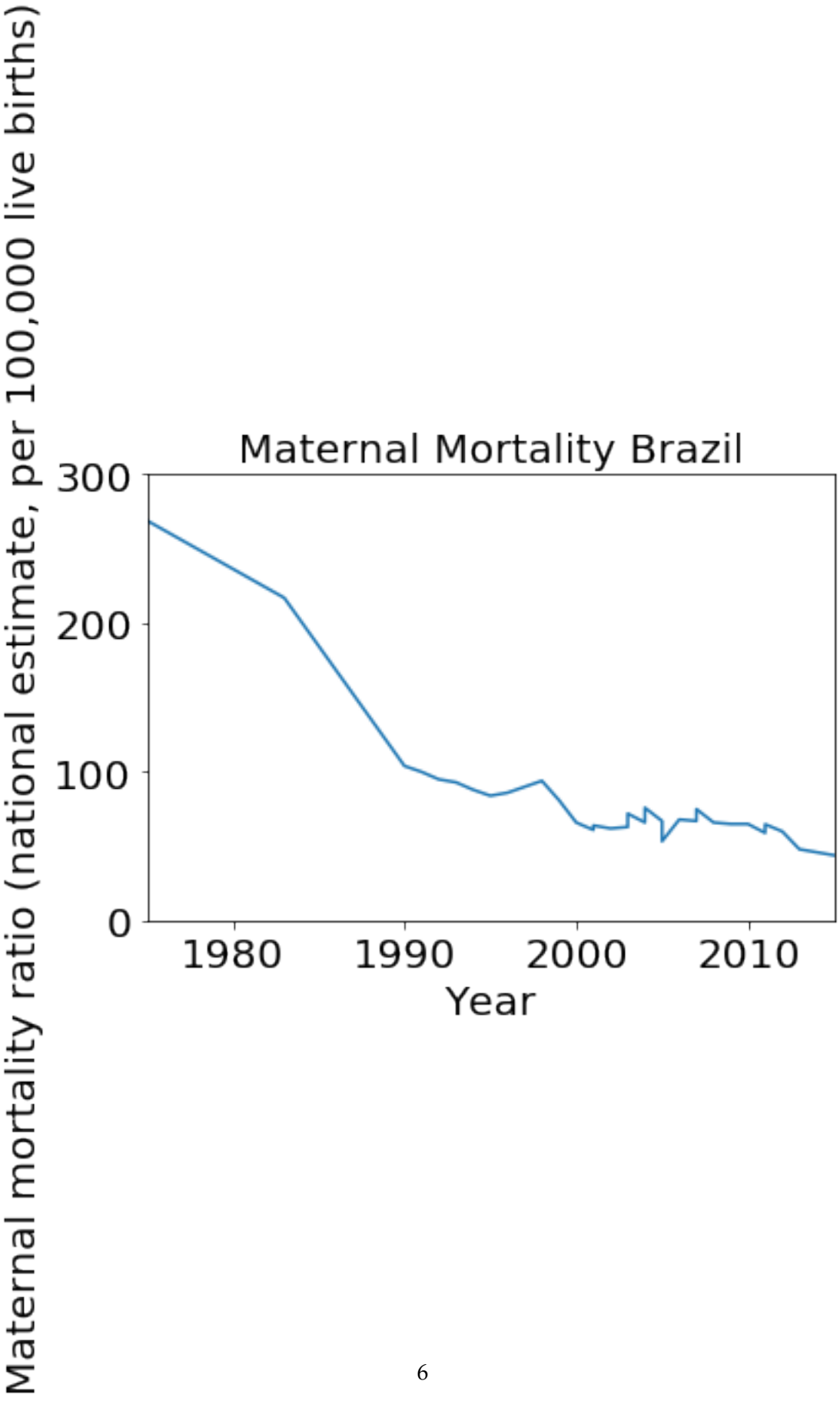
Out[5]:
```
((54, 6), (13, 6), (34, 6), (15, 6))
```

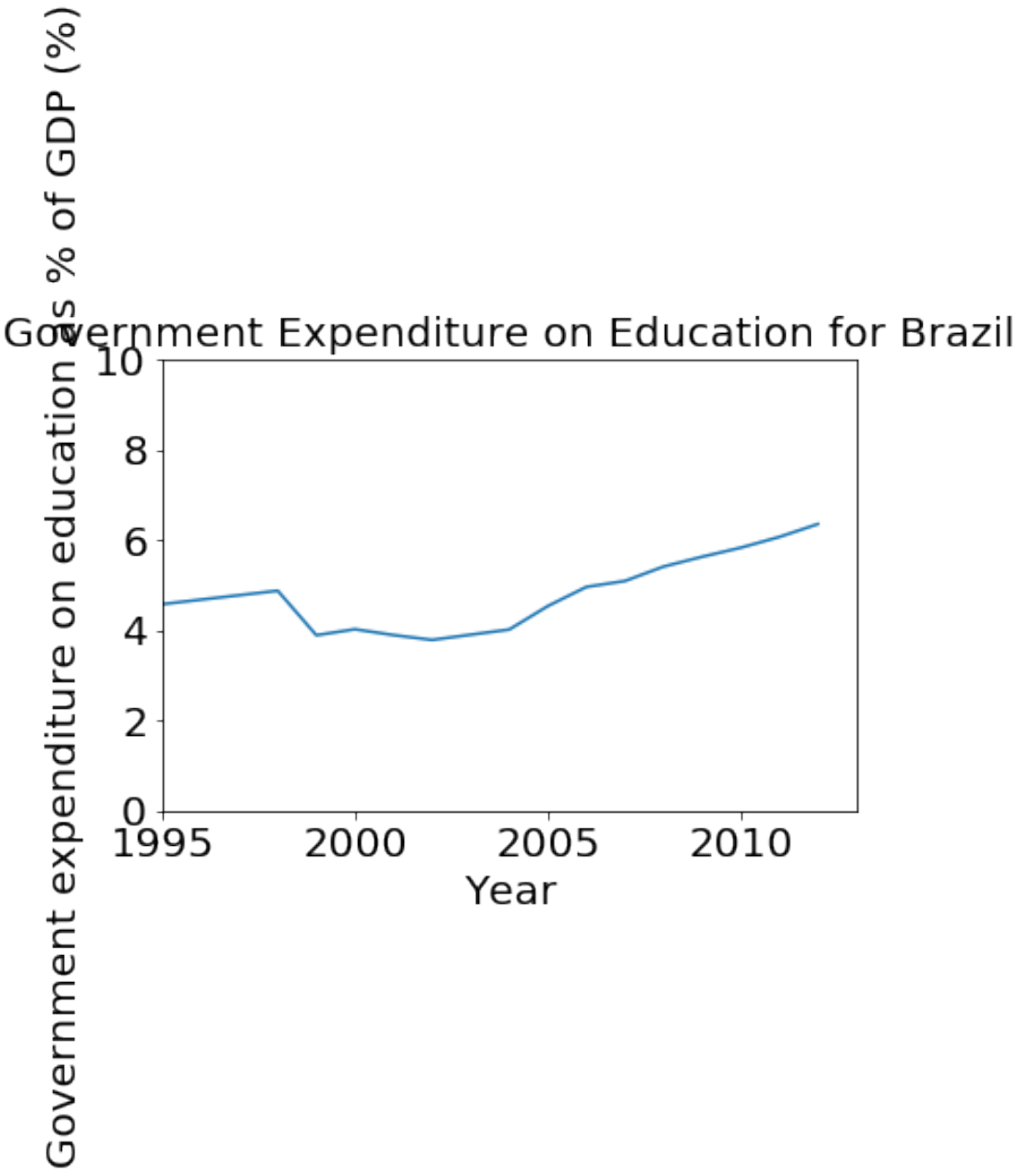
In [6]:
```
# line plot 1
#plt.figure(figsize=(15,10))
plt.figure()
plt.plot(dataBrDeath['Year'].values, dataBrDeath['Value'].values)

# Label the axes
plt.xlabel('Year', fontsize=20)
plt.ylabel(dataBrDeath['IndicatorName'].iloc[0],fontsize=20)

#label the figure
plt.title('Death Rate Brazil',fontsize=20)
plt.tick_params(labelsize=20)

plt.axis([1960, 2013,0,14])
#plt.savefig('line_DeathBR.jpg', bbox_inches='tight')
plt.show()
################################################################################

# line plot 2
#plt.figure(figsize=(15,10))
plt.figure()
plt.plot(dataBrHosp['Year'].values, dataBrHosp['Value'].values)

# Label the axes
plt.xlabel('Year', fontsize=20)
plt.ylabel(dataBrHosp['IndicatorName'].iloc[0],fontsize=20)

#label the figure
plt.title('Hospital Beds Brazil',fontsize=20)
plt.tick_params(labelsize=20)

plt.axis([1960, 2012,0,8])
#plt.savefig('line_HospBR.jpg', bbox_inches='tight')
plt.show()
################################################################################

# line plot 3
#plt.figure(figsize=(15,10))
plt.figure()
plt.plot(dataBrMaternalMort['Year'].values, dataBrMaternalMort['Value'].values)

# Label the axes
plt.xlabel('Year', fontsize=20)
plt.ylabel(dataBrMaternalMort['IndicatorName'].iloc[0],fontsize=20)

#label the figure
plt.title('Maternal Mortality Brazil',fontsize=20)
plt.tick_params(labelsize=20)
```

```
plt.axis([1975, 2015, 0, 300])
#plt.savefig('line_MaternalBR.jpg', bbox_inches='tight')
plt.show()
##############################################################################

# line plot 4
#plt.figure(figsize=(15,10))
plt.figure()
plt.plot(dataBrGovExpEdu['Year'].values, dataBrGovExpEdu['Value'].values)

# Label the axes
plt.xlabel('Year', fontsize=20)
plt.ylabel(dataBrGovExpEdu['IndicatorName'].iloc[0], fontsize=20)

#label the figure
plt.title('Government Expenditure on Education for Brazil', fontsize=20)
plt.tick_params(labelsize=20)

plt.axis([1995, 2013, 0, 10])
#plt.savefig('line_GovExpBR.jpg', bbox_inches='tight')
plt.show()
```


Death Rate Brazil
Death rate, crude (per 1,000 people)
10
5
0
1960
1980
2000
Year


Hospital Beds Brazil
Hospital beds (per 1,000 people)
0
2
4
6
8
1960
1980
2000
Year


Maternal Mortality Brazil
Maternal mortality ratio (national estimate, per 100,000 live births)
300
200
100
0
1980
1990
2000
2010
Year



In [7]:
```
# select Death Rate, Hospital Beds, Maternal Mortality and Government expenditure on e
# indicators matching the ARG for country code and the choosed indicators over time.

argentina = data['CountryCode'].str.contains('ARG')
dataArgDeath = data[argentina & indicatorDeathRate]
dataArgHosp = data[argentina & indicatorHospitalBeds]
```

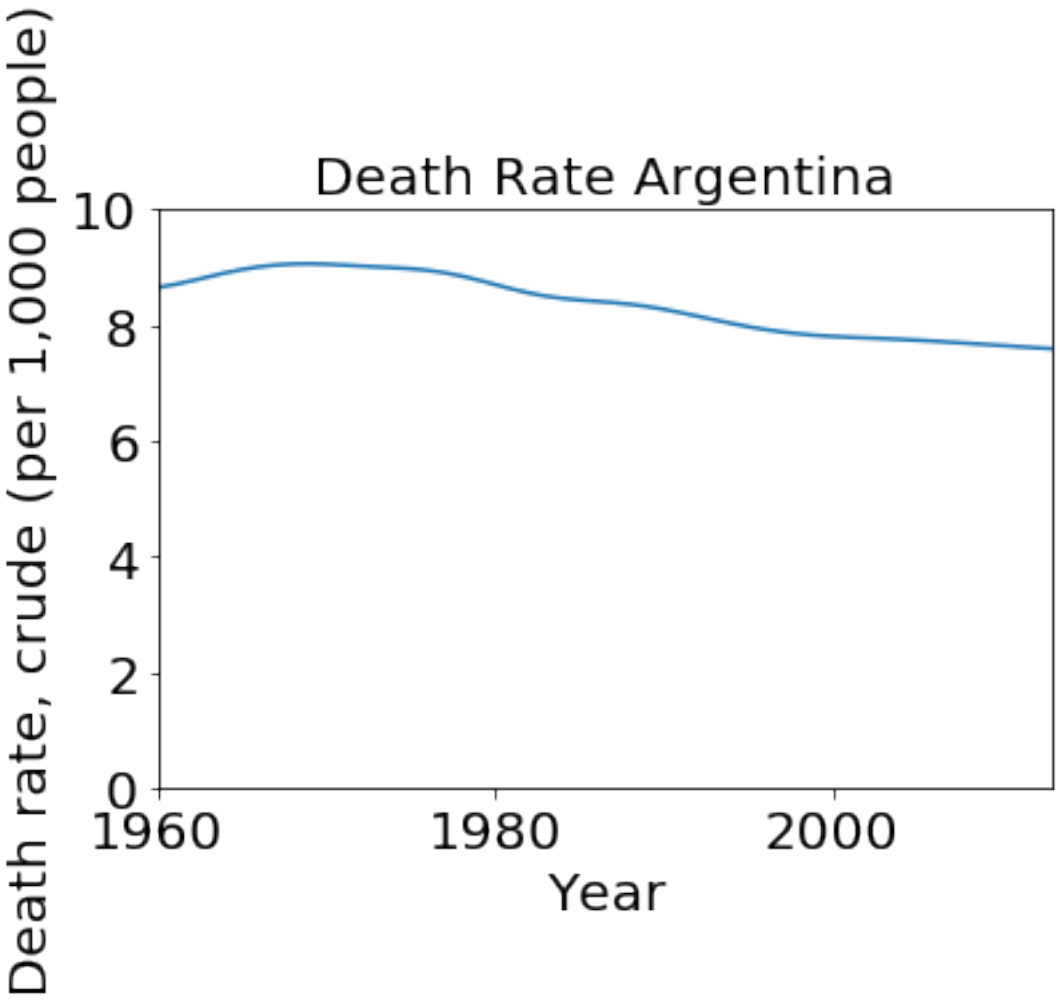
```
dataArgMaternalMort = data[argentina & indicatorMaternalMortality]
#The maternal mortality has two indicators code, so let's keep the same that bra use
dataArgMaternalMort_unique = dataArgMaternalMort[dataArgMaternalMort.IndicatorCode !=
dataArgGovExpEdu = data[argentina & indicatorGovExpEdu]
```

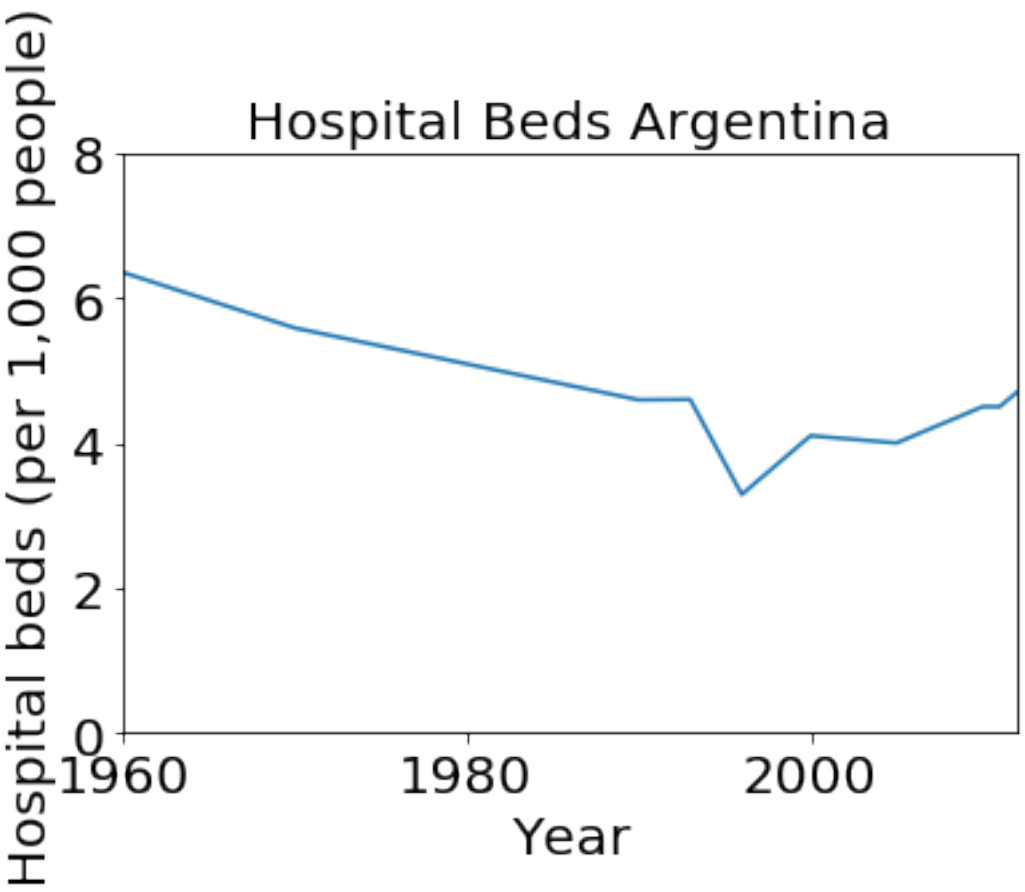
In [8]:
```
dataArgDeath.shape, dataArgHosp.shape, dataArgMaternalMort_unique.shape, dataArgGovExpl
```

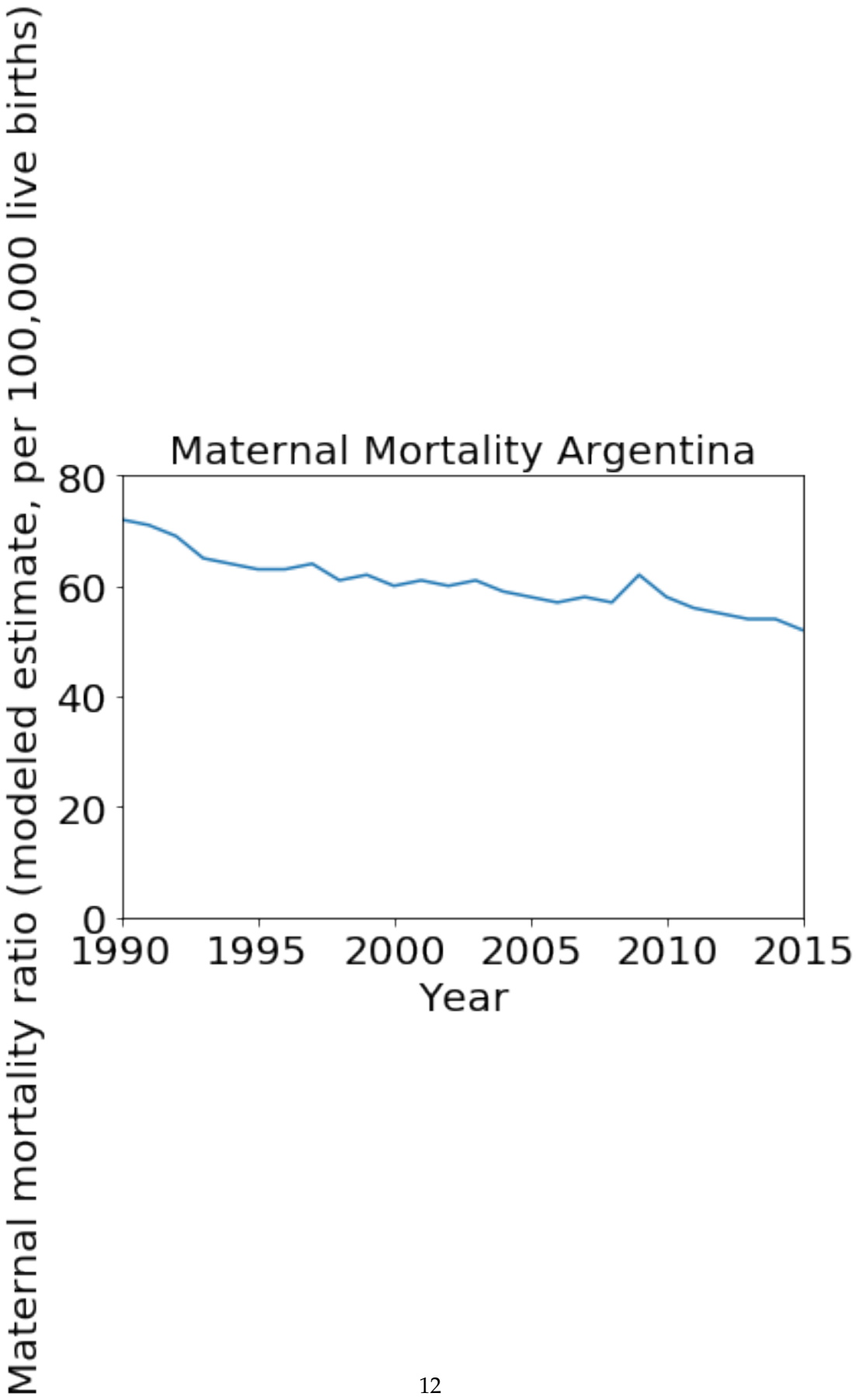
Out[8]:
```
((54, 6), (10, 6), (26, 6), (34, 6))
```

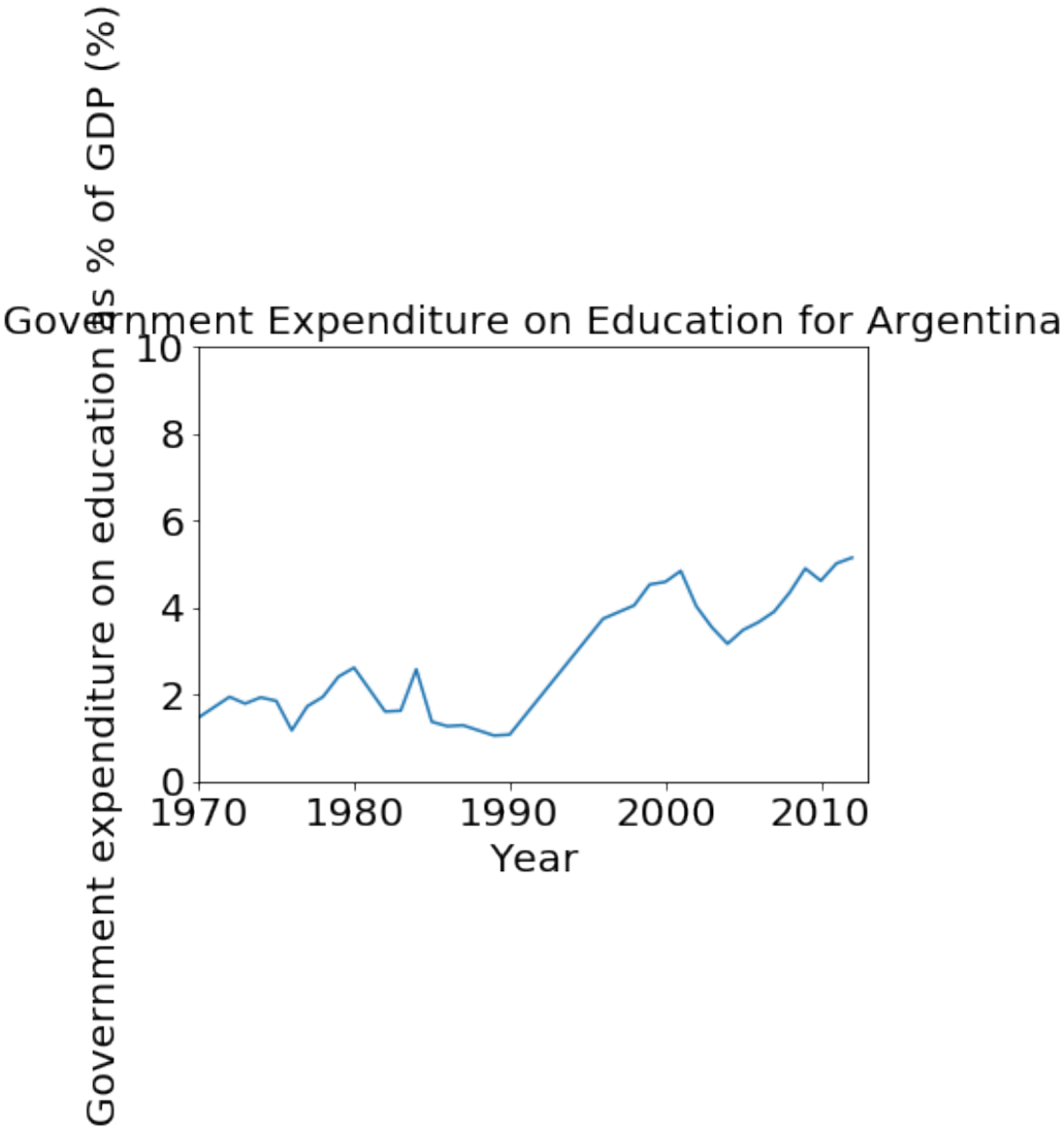
In [9]:
```
# line plot 1
#plt.figure(figsize=(15,10))
plt.figure()
plt.plot(dataArgDeath['Year'].values, dataArgDeath['Value'].values)

# Label the axes
plt.xlabel('Year',fontsize=20)
plt.ylabel(dataArgDeath['IndicatorName'].iloc[0],fontsize=20)

#label the figure
plt.title('Death Rate Argentina',fontsize=20)
plt.tick_params(labelsize=20)

plt.axis([1960, 2013,0,10])
#plt.savefig('line_DeathARG.jpg', bbox_inches='tight')
plt.show()
################################################################################

# line plot 2
#plt.figure(figsize=(15,10))
plt.figure()
plt.plot(dataArgHosp['Year'].values, dataArgHosp['Value'].values)

# Label the axes
plt.xlabel('Year',fontsize=20)
plt.ylabel(dataArgHosp['IndicatorName'].iloc[0],fontsize=20)

#label the figure
plt.title('Hospital Beds Argentina',fontsize=20)
plt.tick_params(labelsize=20)

plt.axis([1960, 2012,0,8])
#plt.savefig('line_HospARG.jpg', bbox_inches='tight')
plt.show()
################################################################################

# line plot 3
#plt.figure(figsize=(15,10))
plt.figure()
```

```
plt.plot(dataArgMaternalMort_unique['Year'].values, dataArgMaternalMort_unique['Value']

# Label the axes
plt.xlabel('Year',fontsize=20)
plt.ylabel(dataArgMaternalMort_unique['IndicatorName'].iloc[0],fontsize=20)

#label the figure
plt.title('Maternal Mortality Argentina',fontsize=20)
plt.tick_params(labelsize=20)

plt.axis([1990, 2015, 0, 80])
#plt.savefig('line_MaternalARG.jpg', bbox_inches='tight')
plt.show()
##############################################################################

# line plot 4
#plt.figure(figsize=(15,10))
plt.figure()
plt.plot(dataArgGovExpEdu['Year'].values, dataArgGovExpEdu['Value'].values)

# Label the axes
plt.xlabel('Year',fontsize=20)
plt.ylabel(dataArgGovExpEdu['IndicatorName'].iloc[0],fontsize=20)

#label the figure
plt.title('Government Expenditure on Education for Argentina',fontsize=20)
plt.tick_params(labelsize=20)

plt.axis([1970, 2013, 0, 10])
#plt.savefig('line_GovExpARG.jpg', bbox_inches='tight')
plt.show()
```


Death Rate Argentina
Death rate, crude (per 1,000 people)
10
8
6
4
2
0
1960
1980
2000
Year


Hospital Beds Argentina
Hospital beds (per 1,000 people)
0
2
4
6
8
1960
1980
2000
Year


Maternal Mortality Argentina
Maternal mortality ratio (modeled estimate, per 100,000 live births)
0
20
40
60
80
1990
1995
2000
2005
2010
2015
Year



0.1 Deleting unnecessary columns and rename the Value to "Country Value Indicator"

```
In [10]: del dataBrDeath['CountryName']
         del dataBrDeath['CountryCode']
         del dataBrDeath['IndicatorName']
         del dataBrDeath['IndicatorCode']
```

```
In [11]: dataBrDeath_simple = dataBrDeath.rename(columns={'Value': 'BR Value Death rate'})
         dataBrDeath_simple.head()
```

Out[11]:

	Year	BR Value Death rate
5649	1960	13.410
29730	1961	13.093
56767	1962	12.762
85155	1963	12.419
113867	1964	12.066

In [12]:
```
del dataBrHosp['CountryName']
del dataBrHosp['CountryCode']
del dataBrHosp['IndicatorName']
del dataBrHosp['IndicatorCode']
```

In [13]:
```
dataBrHosp_simple = dataBrHosp.rename(columns={'Value': 'BR Value Hospital Beds'})
dataBrHosp_simple.head()
```

Out[13]:

	Year	BR Value Hospital Beds
5713	1960	3.201795
312182	1970	3.691800
654688	1975	4.995100
1628982	1987	3.560300
1907479	1990	3.345300

In [14]:
```
del dataBrMaternalMort['CountryName']
del dataBrMaternalMort['CountryCode']
del dataBrMaternalMort['IndicatorName']
del dataBrMaternalMort['IndicatorCode']
```

In [15]:
```
dataBrMaternalMort_simple = dataBrMaternalMort.rename(columns={'Value': 'BR Value Mat
dataBrMaternalMort_simple.head()
```

Out[15]:

	Year	BR Value Maternal Mortality
654745	1975	269.0
1281380	1983	217.0
1907588	1990	104.0
2024764	1991	100.0
2141642	1992	95.0

In [16]:
```
del dataBrGovExpEdu['CountryName']
del dataBrGovExpEdu['CountryCode']
del dataBrGovExpEdu['IndicatorName']
del dataBrGovExpEdu['IndicatorCode']
```

In [17]:
```
dataBrGovExpEdu_simple = dataBrGovExpEdu.rename(columns={'Value': 'BR Value Gov Expen
dataBrGovExpEdu_simple.head()
```

Out[17]:

	Year	BR Value Gov Expend Educat
2515806	1995	4.57043
2922767	1998	4.86875
3061028	1999	3.88167
3206545	2000	4.01458
3359890	2001	3.88489

```
In [18]: del dataArgDeath['CountryName']
         del dataArgDeath['CountryCode']
         del dataArgDeath['IndicatorName']
         del dataArgDeath['IndicatorCode']

In [19]: dataArgDeath_simple = dataArgDeath.rename(columns={'Value': 'ARG Value Death rate'})
         dataArgDeath_simple.head()

Out[19]:         Year  ARG Value Death rate
         3982    1960                 8.637
         27803   1961                 8.689
         54669   1962                 8.752
         83049   1963                 8.821
         111740  1964                 8.889

In [20]: del dataArgHosp['CountryName']
         del dataArgHosp['CountryCode']
         del dataArgHosp['IndicatorName']
         del dataArgHosp['IndicatorCode']

In [21]: dataArgHosp_simple = dataArgHosp.rename(columns={'Value': 'ARG Value Hospital Beds'})
         dataArgHosp_simple.head()

Out[21]:          Year  ARG Value Hospital Beds
         4037     1960                 6.352251
         308216   1970                 5.585800
         1898770  1990                 4.594300
         2254092  1993                 4.600000
         2640017  1996                 3.290000

In [22]: del dataArgMaternalMort_unique['CountryName']
         del dataArgMaternalMort_unique['CountryCode']
         del dataArgMaternalMort_unique['IndicatorName']
         del dataArgMaternalMort_unique['IndicatorCode']

In [23]: dataArgMaternalMort_unique_simple = dataArgMaternalMort_unique.rename(columns={'Value
         dataArgMaternalMort_unique_simple.head()

Out[23]:          Year  ARG Value Maternal Mortality
         1898845  1990                          72.0
         2016017  1991                          71.0
         2132515  1992                          69.0
         2254178  1993                          65.0
         2377978  1994                          64.0

In [24]: del dataArgGovExpEdu['CountryName']
         del dataArgGovExpEdu['CountryCode']
         del dataArgGovExpEdu['IndicatorName']
         del dataArgGovExpEdu['IndicatorCode']
```

```
In [25]: dataArgGovExpEdu_simple = dataArgGovExpEdu.rename(columns={'Value': 'ARG Value Gov Exp
         dataArgGovExpEdu_simple.head()
```

Out[25]:

	Year	ARG Value Gov Expend Educat
308167	1970	1.45809
439810	1972	1.93620
509690	1973	1.78052
579274	1974	1.92352
649763	1975	1.84360

0.2 ScatterPlot for comparing Death Rate against Hospital Beds and Maternal Mortality against Gov Expend Educat

First, we'll need to make sure we're looking at the same time frames

0.3 Brazil

```
In [26]: print("   Death Rate BRA: min =", dataBrDeath_simple['Year'].min(), "max:", dataBrDeat
         print("Hospital Beds BRA: min =", dataBrHosp_simple['Year'].min(), "max:", dataBrHosp_
```

```
   Death Rate BRA: min = 1960 max: 2013
Hospital Beds BRA: min = 1960 max: 2012
```

```
In [27]: #Merge by year
         dataBrDeath_Hosp = dataBrDeath_simple.merge(dataBrHosp_simple, on='Year', how='inner')
         dataBrDeath_Hosp.head()
```

Out[27]:

	Year	BR Value Death rate	BR Value Hospital Beds
0	1960	13.410	3.201795
1	1970	10.081	3.691800
2	1975	9.197	4.995100
3	1987	7.606	3.560300
4	1990	7.175	3.345300

```
In [28]: print("Maternal Mortality BRA: min =", dataBrMaternalMort_simple['Year'].min(), "max:"
         print(" Gov Expend Educat BRA: min =", dataBrGovExpEdu_simple['Year'].min(), "max:", d
```

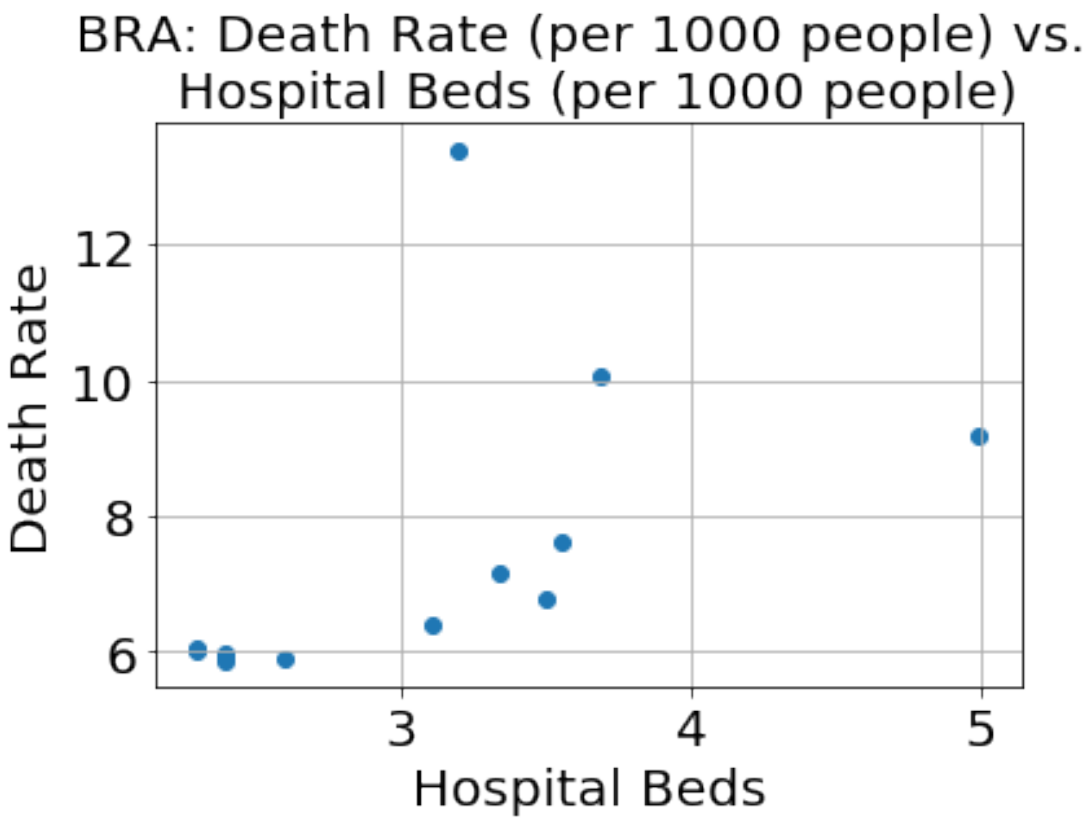
```
Maternal Mortality BRA: min = 1975 max: 2015
 Gov Expend Educat BRA: min = 1995 max: 2012
```

```
In [29]: #Merge by year
         dataBrMaternal_GovExp = dataBrMaternalMort_simple.merge(dataBrGovExpEdu_simple, on='Ye
         dataBrMaternal_GovExp.head()
```

Out[29]:

	Year	BR Value Maternal Mortality	BR Value Gov Expend Educat
0	1995	84.0	4.57043
1	1998	94.0	4.86875
2	1999	81.0	3.88167
3	2000	66.0	4.01458
4	2001	61.0	3.88489

In [30]:
```
#fig, axis = plt.subplots(figsize=(15,10))
fig, axis = plt.subplots()
# Grid lines, Xticks, Xlabel, Ylabel

axis.yaxis.grid(True)
axis.xaxis.grid(True)
axis.set_title('BRA: Death Rate (per 1000 people) vs. \n Hospital Beds (per 1000 peopl
axis.set_xlabel('Hospital Beds' ,fontsize=20)
axis.set_ylabel('Death Rate' ,fontsize=20)

XBRdeath = dataBrDeath_Hosp['BR Value Hospital Beds']
YBRHosp = dataBrDeath_Hosp['BR Value Death rate']
plt.tick_params(labelsize=20)

axis.scatter(XBRdeath, YBRHosp)
#plt.savefig('scatBR_DeathvsHosp.jpg', bbox_inches='tight')
plt.show()
```

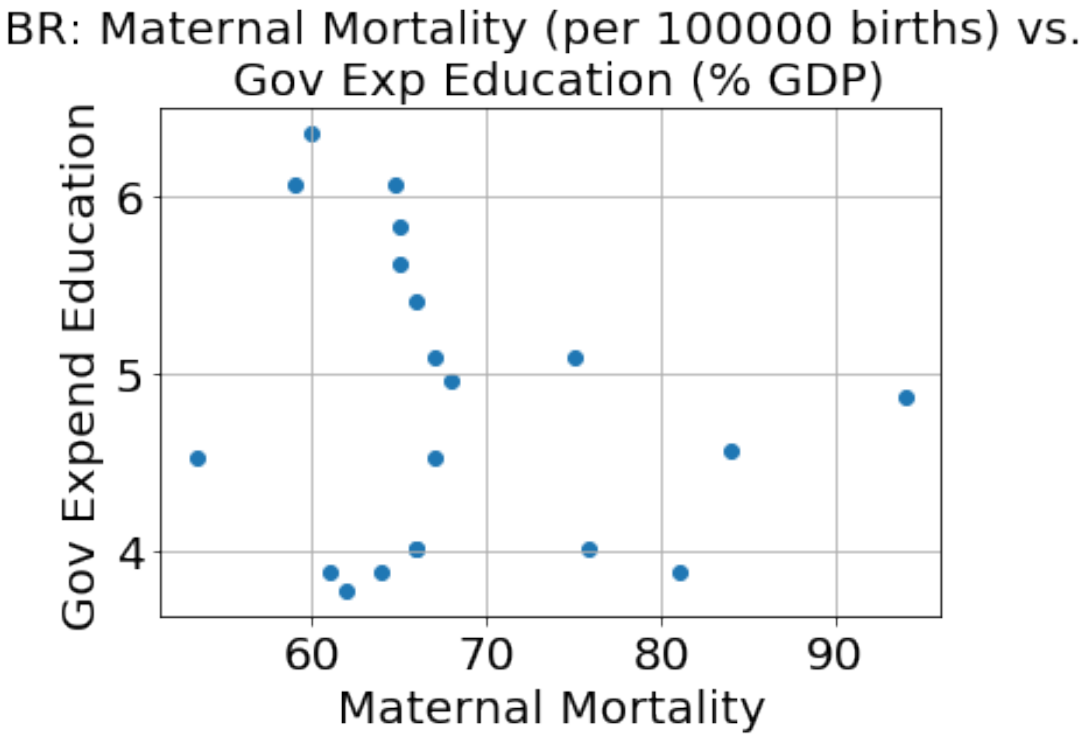


Let's test this by looking at correlation.

In [31]:
```
np.corrcoef(XBRdeath, YBRHosp)
```

```
Out[31]: array([[1.        , 0.54145374],
                [0.54145374, 1.        ]])
```

```
In [32]: #fig, axis = plt.subplots(figsize=(15,10))
         fig, axis = plt.subplots()
         # Grid lines, Xticks, Xlabel, Ylabel

         axis.yaxis.grid(True)
         axis.xaxis.grid(True)
         axis.set_title('BR: Maternal Mortality (per 100000 births) vs. \n Gov Exp Education (
         axis.set_xlabel('Maternal Mortality' ,fontsize=20)
         axis.set_ylabel('Gov Expend Education' ,fontsize=20)

         XBRMaternal = dataBrMaternal_GovExp['BR Value Maternal Mortality']
         YBRGov = dataBrMaternal_GovExp['BR Value Gov Expend Educat']
         plt.tick_params(labelsize=20)

         axis.scatter(XBRMaternal, YBRGov)
         #plt.savefig('scatBR_MatvsGov.jpg', bbox_inches='tight')
         plt.show()
```



Let's test this by looking at correlation.

```
In [33]: np.corrcoef(XBRMaternal, YBRGov)
```

```
Out[33]: array([[ 1.        , -0.19319095],
                [-0.19319095,  1.        ]])
```

Let's see the correlations with each other indicators

```
In [34]: #Corr with Death Rate and Maternal Mortality
         #Merge by year
         dataBrDeath_Maternal = dataBrDeath_simple.merge(dataBrMaternalMort_simple, on='Year',
         print(dataBrDeath_Maternal.head())
         np.corrcoef(dataBrDeath_Maternal['BR Value Death rate'], dataBrDeath_Maternal['BR Valu
```

```
   Year  BR Value Death rate  BR Value Maternal Mortality
0  1975                9.197                         269.0
1  1983                8.242                         217.0
2  1990                7.175                         104.0
3  1991                7.042                         100.0
4  1992                6.909                          95.0
```

```
Out[34]: array([[1.        , 0.95227017],
                [0.95227017, 1.        ]])
```

```
In [35]: #Corr with Death Rate and Gov Expend Educt
         #Merge by year
         dataBrDeath_GovExp = dataBrDeath_simple.merge(dataBrGovExpEdu_simple, on='Year', how=
         print(dataBrDeath_GovExp.head())
         np.corrcoef(dataBrDeath_GovExp['BR Value Death rate'], dataBrDeath_GovExp['BR Value G
```

```
   Year  BR Value Death rate  BR Value Gov Expend Educat
0  1995                6.514                     4.57043
1  1998                6.170                     4.86875
2  1999                6.080                     3.88167
3  2000                6.006                     4.01458
4  2001                5.950                     3.88489
```

```
Out[35]: array([[ 1.        , -0.01753263],
                [-0.01753263,  1.        ]])
```

```
In [36]: #Corr with Hospital Beds and Maternal Mortality
         #Merge by year
         dataBrHosp_Maternal = dataBrHosp_simple.merge(dataBrMaternalMort_simple, on='Year', ho
         print(dataBrHosp_Maternal.head())
         np.corrcoef(dataBrHosp_Maternal['BR Value Hospital Beds'], dataBrHosp_Maternal['BR Val
```

```
   Year  BR Value Hospital Beds  BR Value Maternal Mortality
0  1975                  4.9951                         269.0
1  1990                  3.3453                         104.0
2  1993                  3.5000                          93.0
3  1996                  3.1100                          86.0
4  2002                  2.6000                          62.0
```

```
Out[36]: array([[1.        , 0.94512472],
                [0.94512472, 1.        ]])
```

```
In [37]: #Corr with Hospital Beds and Gov Expend Educt
         #Merge by year
         dataBrHosp_GovExp = dataBrHosp_simple.merge(dataBrGovExpEdu_simple, on='Year', how='in
         print(dataBrHosp_GovExp.head())
         np.corrcoef(dataBrHosp_GovExp['BR Value Hospital Beds'], dataBrHosp_GovExp['BR Value
```

```
   Year  BR Value Hospital Beds  BR Value Gov Expend Educat
0  2002                     2.6                     3.77820
1  2005                     2.4                     4.52778
2  2009                     2.4                     5.62147
3  2010                     2.4                     5.82225
4  2011                     2.3                     6.06058
```

```
Out[37]: array([[ 1.        , -0.89090677],
                [-0.89090677,  1.        ]])
```

0.3.1 There are strange correlation, let's see if there is relation with the Total population

```
In [38]: indicatorPopTot = data['IndicatorName'].str.contains("Population, total")

         # indicators matching the BRA for country code and the choosed indicators over time.
         dataBrPopTot = data[brazil & indicatorPopTot]
```

```
In [39]: dataBrPopTot.shape
```

```
Out[39]: (55, 6)
```

```
In [40]: del dataBrPopTot['CountryName']
         del dataBrPopTot['CountryCode']
         del dataBrPopTot['IndicatorName']
         del dataBrPopTot['IndicatorCode']
```

```
In [41]: dataBrPopTot_simple = dataBrPopTot.rename(columns={'Value': 'BRA Value Total Populatio
         dataBrPopTot_simple.head()
```

```
Out[41]:         Year  BRA Value Total Population
         5800    1960                  72493585.0
         29893   1961                  74706888.0
         56944   1962                  77007549.0
         85331   1963                  79368453.0
         114044  1964                  81751802.0
```

```
In [42]: dataBrPop_Death = dataBrPopTot_simple.merge(dataBrDeath_simple, on='Year', how='inner
         print(dataBrPop_Death.head())
         np.corrcoef(dataBrPop_Death['BRA Value Total Population'], dataBrPop_Death['BR Value
```

```
   Year  BRA Value Total Population  BR Value Death rate
0  1960                  72493585.0                13.410
1  1961                  74706888.0                13.093
2  1962                  77007549.0                12.762
3  1963                  79368453.0                12.419
4  1964                  81751802.0                12.066
```

```
Out[42]: array([[ 1.        , -0.95630809],
                [-0.95630809,  1.        ]])
```

```
In [43]: dataBrPop_Hosp = dataBrPopTot_simple.merge(dataBrHosp_simple, on='Year', how='inner')
         print(dataBrPop_Hosp.head())
         np.corrcoef(dataBrPop_Hosp['BRA Value Total Population'], dataBrPop_Hosp['BR Value Hos
```

```
   Year  BRA Value Total Population  BR Value Hospital Beds
0  1960                  72493585.0                 3.201795
1  1970                  95982453.0                 3.691800
2  1975                 108431284.0                 4.995100
3  1987                 142437479.0                 3.560300
4  1990                 150393143.0                 3.345300
```

```
Out[43]: array([[ 1.        , -0.75427574],
                [-0.75427574,  1.        ]])
```

```
In [44]: dataBrPop_Mort = dataBrPopTot_simple.merge(dataBrMaternalMort_simple, on='Year', how=
         print(dataBrPop_Mort.head())
         np.corrcoef(dataBrPop_Mort['BRA Value Total Population'], dataBrPop_Mort['BR Value Mat
```

```
   Year  BRA Value Total Population  BR Value Maternal Mortality
0  1975                 108431284.0                         269.0
1  1983                 131014337.0                         217.0
2  1990                 150393143.0                         104.0
3  1991                 152916852.0                         100.0
4  1992                 155379009.0                          95.0
```

```
Out[44]: array([[ 1.        , -0.86443462],
                [-0.86443462,  1.        ]])
```

```
In [45]: dataBrPop_Gov = dataBrPopTot_simple.merge(dataBrGovExpEdu_simple, on='Year', how='inne
         print(dataBrPop_Gov.head())
         np.corrcoef(dataBrPop_Gov['BRA Value Total Population'], dataBrPop_Gov['BR Value Gov E
```

```
   Year  BRA Value Total Population  BR Value Gov Expend Educat
0  1995                 162755054.0                      4.57043
1  1998                 170516482.0                      4.86875
2  1999                 173153066.0                      3.88167
3  2000                 175786441.0                      4.01458
4  2001                 178419396.0                      3.88489
```

```
Out[45]: array([[1.        , 0.75184343],
                [0.75184343, 1.        ]])
```

0.4 Argentina

```
In [46]: print("   Death Rate ARG: min =", dataArgDeath_simple['Year'].min(), "max:", dataArgDe
         print("Hospital Beds ARG: min =", dataArgHosp_simple['Year'].min(), "max:", dataArgHos
```

```
   Death Rate ARG: min = 1960 max: 2013
Hospital Beds ARG: min = 1960 max: 2012
```

```
In [47]: #Merge by year
         dataArgDeath_Hosp = dataArgDeath_simple.merge(dataArgHosp_simple, on='Year', how='inne
         dataArgDeath_Hosp.head()
```

```
Out[47]:    Year  ARG Value Death rate  ARG Value Hospital Beds
         0  1960                 8.637                 6.352251
         1  1970                 9.039                 5.585800
         2  1990                 8.264                 4.594300
         3  1993                 8.082                 4.600000
         4  1996                 7.916                 3.290000
```

```
In [48]: print("Maternal Mortality ARG: min =", dataArgMaternalMort_unique_simple['Year'].min()
         print(" Gov Expend Educat ARG: min =", dataArgGovExpEdu_simple['Year'].min(), "max:",
```

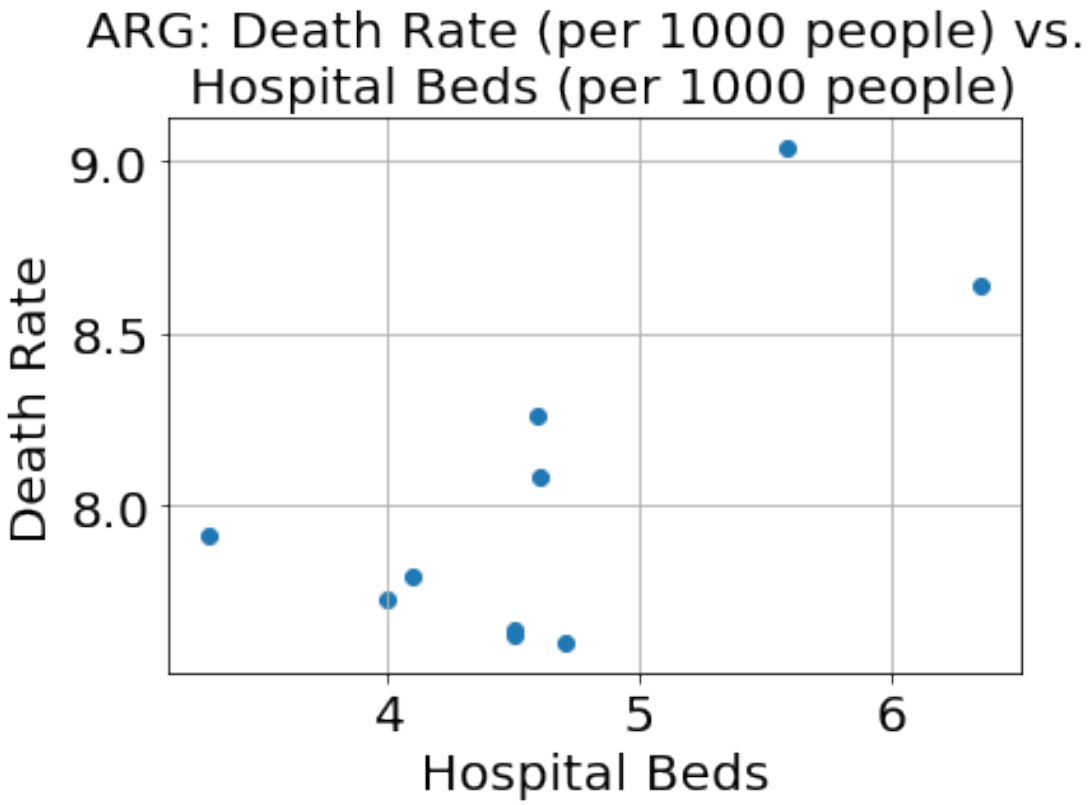
```
Maternal Mortality ARG: min = 1990 max: 2015
 Gov Expend Educat ARG: min = 1970 max: 2012
```

```
In [49]: #Merge by year
         dataArgMaternal_GovExp = dataArgMaternalMort_unique_simple.merge(dataArgGovExpEdu_simp
         dataArgMaternal_GovExp.head()
```

```
Out[49]:    Year  ARG Value Maternal Mortality  ARG Value Gov Expend Educat
         0  1990                          72.0                       1.06738
         1  1996                          63.0                       3.73198
         2  1998                          61.0                       4.03987
         3  1999                          62.0                       4.52168
         4  2000                          60.0                       4.58031
```

```
In [50]: #fig, axis = plt.subplots(figsize=(15,10))
         fig, axis = plt.subplots()
         # Grid lines, Xticks, Xlabel, Ylabel

         axis.yaxis.grid(True)
         axis.xaxis.grid(True)
         axis.set_title('ARG: Death Rate (per 1000 people) vs. \n Hospital Beds (per 1000 peopl
         axis.set_xlabel('Hospital Beds' ,fontsize=20)
```

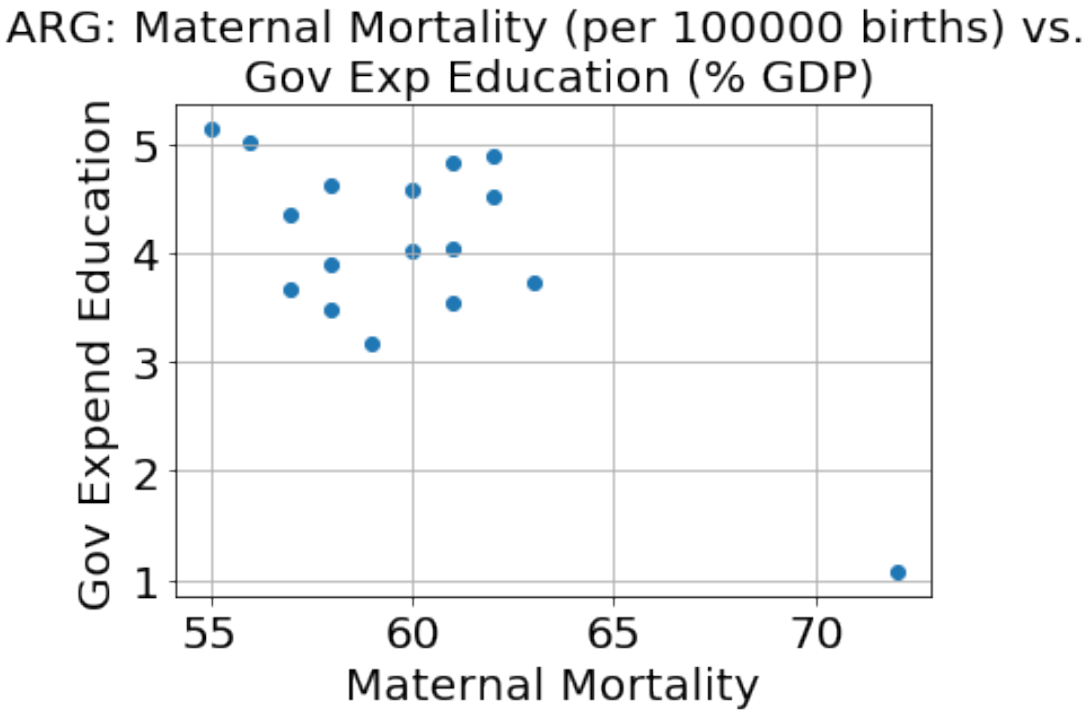
```
axis.set_ylabel('Death Rate' ,fontsize=20)
plt.tick_params(labelsize=20)

XARGdeath = dataArgDeath_Hosp['ARG Value Hospital Beds']
YARGHosp = dataArgDeath_Hosp['ARG Value Death rate']

axis.scatter(XARGdeath, YARGHosp)
#plt.savefig('scatARG_DeathvsHosp.jpg', bbox_inches='tight')
plt.show()
```



Let's test this by looking at correlation.

```
In [51]: np.corrcoef(XARGdeath, YARGHosp)

Out[51]: array([[1.        , 0.69237676],
                [0.69237676, 1.        ]])

In [52]: #fig, axis = plt.subplots(figsize=(15,10))
         fig, axis = plt.subplots()
         # Grid lines, Xticks, Xlabel, Ylabel

         axis.yaxis.grid(True)
         axis.xaxis.grid(True)
```

```
axis.set_title('ARG: Maternal Mortality (per 100000 births) vs. \n Gov Exp Education
axis.set_xlabel('Maternal Mortality' ,fontsize=20)
axis.set_ylabel('Gov Expend Education' ,fontsize=20)
plt.tick_params(labelsize=20)

XARGMaternal = dataArgMaternal_GovExp['ARG Value Maternal Mortality']
YARGGov = dataArgMaternal_GovExp['ARG Value Gov Expend Educat']

axis.scatter(XARGMaternal, YARGGov)
#plt.savefig('scatARG_MatvsGov.jpg', bbox_inches='tight')
plt.show()
```



Let's test this by looking at correlation.

```
In [53]: np.corrcoef(XARGMaternal, YARGGov)

Out[53]: array([[ 1.        , -0.70142767],
                [-0.70142767,  1.        ]])

In [54]: #Corr with Death Rate and Maternal Mortality
         #Merge by year
         dataArgDeath_Maternal = dataArgDeath_simple.merge(dataArgMaternalMort_unique_simple, 
         print(dataArgDeath_Maternal.head())
         np.corrcoef(dataArgDeath_Maternal['ARG Value Death rate'], dataArgDeath_Maternal['ARG
```

```
   Year  ARG Value Death rate  ARG Value Maternal Mortality
0  1990                 8.264                          72.0
1  1991                 8.207                          71.0
2  1992                 8.145                          69.0
3  1993                 8.082                          65.0
4  1994                 8.021                          64.0
```

```
Out[54]: array([[1.        , 0.94970638],
                [0.94970638, 1.        ]])
```

```
In [55]: #Corr with Death Rate and Gov Expend Educt
         #Merge by year
         dataArgDeath_GovExp = dataArgDeath_simple.merge(dataArgGovExpEdu_simple, on='Year', ho
         print(dataArgDeath_GovExp.head())
         np.corrcoef(dataArgDeath_GovExp['ARG Value Death rate'], dataArgDeath_GovExp['ARG Valu
```

```
   Year  ARG Value Death rate  ARG Value Gov Expend Educat
0  1970                 9.039                      1.45809
1  1972                 9.008                      1.93620
2  1973                 8.993                      1.78052
3  1974                 8.979                      1.92352
4  1975                 8.961                      1.84360
```

```
Out[55]: array([[ 1.       , -0.8299004],
                [-0.8299004,  1.       ]])
```

```
In [56]: #Corr with Hospital Beds and Maternal Mortality
         #Merge by year
         dataArgHosp_Maternal = dataArgHosp_simple.merge(dataArgMaternalMort_unique_simple, on=
         print(dataArgHosp_Maternal.head())
         np.corrcoef(dataArgHosp_Maternal['ARG Value Hospital Beds'], dataArgHosp_Maternal['ARG
```

```
   Year  ARG Value Hospital Beds  ARG Value Maternal Mortality
0  1990                   4.5943                          72.0
1  1993                   4.6000                          65.0
2  1996                   3.2900                          63.0
3  2000                   4.1000                          60.0
4  2005                   4.0000                          58.0
```

```
Out[56]: array([[ 1.        , -0.02674022],
                [-0.02674022,  1.        ]])
```

```
In [57]: #Corr with Hospital Beds and Gov Expend Educt
         #Merge by year
         dataArgHosp_GovExp = dataArgHosp_simple.merge(dataArgGovExpEdu_simple, on='Year', how=
         print(dataArgHosp_GovExp.head())
         np.corrcoef(dataArgHosp_GovExp['ARG Value Hospital Beds'], dataArgHosp_GovExp['ARG Val
```

```
   Year  ARG Value Hospital Beds  ARG Value Gov Expend Educat
0  1970                   5.5858                      1.45809
1  1990                   4.5943                      1.06738
2  1996                   3.2900                      3.73198
3  2000                   4.1000                      4.58031
4  2005                   4.0000                      3.47405
```

```
Out[57]: array([[ 1.       , -0.3751571],
                [-0.3751571,  1.       ]])
```

0.4.1 There are strange correlation, let's see if there is relation with the Total population

```
In [58]: # indicators matching the ARG for country code and the choosed indicators over time.
         dataArgPopTot = data[argentina & indicatorPopTot]

In [59]: dataArgPopTot.shape

Out[59]: (55, 6)

In [60]: del dataArgPopTot['CountryName']
         del dataArgPopTot['CountryCode']
         del dataArgPopTot['IndicatorName']
         del dataArgPopTot['IndicatorCode']

In [61]: dataArgPopTot_simple = dataArgPopTot.rename(columns={'Value': 'ARG Value Total Popula
         dataArgPopTot_simple.head()

Out[61]:         Year  ARG Value Total Population
         4104    1960                  20619075.0
         27934   1961                  20953079.0
         54830   1962                  21287682.0
         83213   1963                  21621845.0
         111905  1964                  21953926.0

In [62]: dataArgPop_Death = dataArgPopTot_simple.merge(dataArgDeath_simple, on='Year', how='in
         print(dataArgPop_Death.head())
         np.corrcoef(dataArgPop_Death['ARG Value Total Population'], dataArgPop_Death['ARG Valu
```

```
   Year  ARG Value Total Population  ARG Value Death rate
0  1960                  20619075.0                 8.637
1  1961                  20953079.0                 8.689
2  1962                  21287682.0                 8.752
3  1963                  21621845.0                 8.821
4  1964                  21953926.0                 8.889
```

```
Out[62]: array([[ 1.        , -0.95680045],
                [-0.95680045,  1.        ]])
```

```
In [63]: dataArgPop_Hosp = dataArgPopTot_simple.merge(dataArgHosp_simple, on='Year', how='inner
         print(dataArgPop_Hosp.head())
         np.corrcoef(dataArgPop_Hosp['ARG Value Total Population'], dataArgPop_Hosp['ARG Value
```

	Year	ARG Value Total Population	ARG Value Hospital Beds
0	1960	20619075.0	6.352251
1	1970	23973062.0	5.585800
2	1990	32729740.0	4.594300
3	1993	34110912.0	4.600000
4	1996	35419683.0	3.290000

```
Out[63]: array([[ 1.        , -0.72719188],
                [-0.72719188,  1.        ]])

In [64]: dataArgPop_Mort = dataArgPopTot_simple.merge(dataArgMaternalMort_unique_simple, on='Y
         print(dataArgPop_Mort.head())
         np.corrcoef(dataArgPop_Mort['ARG Value Total Population'], dataArgPop_Mort['ARG Value
```

	Year	ARG Value Total Population	ARG Value Maternal Mortality
0	1990	32729740.0	72.0
1	1991	33193920.0	71.0
2	1992	33655149.0	69.0
3	1993	34110912.0	65.0
4	1994	34558114.0	64.0

```
Out[64]: array([[ 1.        , -0.91768838],
                [-0.91768838,  1.        ]])

In [65]: dataArgPop_Gov = dataArgPopTot_simple.merge(dataArgGovExpEdu_simple, on='Year', how='
         print(dataArgPop_Gov.head())
         np.corrcoef(dataArgPop_Gov['ARG Value Total Population'], dataArgPop_Gov['ARG Value G
```

	Year	ARG Value Total Population	ARG Value Gov Expend Educat
0	1970	23973062.0	1.45809
1	1972	24782950.0	1.93620
2	1973	25213388.0	1.78052
3	1974	25644505.0	1.92352
4	1975	26066975.0	1.84360

```
Out[65]: array([[1.       , 0.8430293],
                [0.8430293, 1.       ]])
```

1 See a histogram with all countries

```
In [66]: # select Death rate for all countries in 2010
         hist_indicator = 'Death rate, crude'
```

```
hist_year = 2010

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

# apply our mask
deathRate_2010 = data[mask1 & mask2]
deathRate_2010.head()
```

Out[66]:

	CountryName	CountryCode	IndicatorName	IndicatorCode	Year	Value
4839128	Arab World	ARB	Death rate, crude (per 1,000 people)	SP.DYN.CDRT.IN	2010	5.845962
4839673	Caribbean small states	CSS	Death rate, crude (per 1,000 people)	SP.DYN.CDRT.IN	2010	7.904228
4840227	Central Europe and the Baltics	CEB	Death rate, crude (per 1,000 people)	SP.DYN.CDRT.IN	2010	11.429917
4840893	East Asia & Pacific (all income levels)	EAS	Death rate, crude (per 1,000 people)	SP.DYN.CDRT.IN	2010	7.121761
4841541	East Asia & Pacific (developing only)	EAP	Death rate, crude (per 1,000 people)	SP.DYN.CDRT.IN	2010	7.048825

For how many countries do we have Death rate data in 2010

In [67]:
```
print(len(deathRate_2010))
```

239

In [68]:
```
deathRate_2010[deathRate_2010['CountryCode'] == 'BRA']
```

Out[68]:

	CountryName	CountryCode	IndicatorName	IndicatorCode	Year	Value
4882538	Brazil	BRA	Death rate, crude (per 1,000 people)	SP.DYN.CDRT.IN	2010	5.97

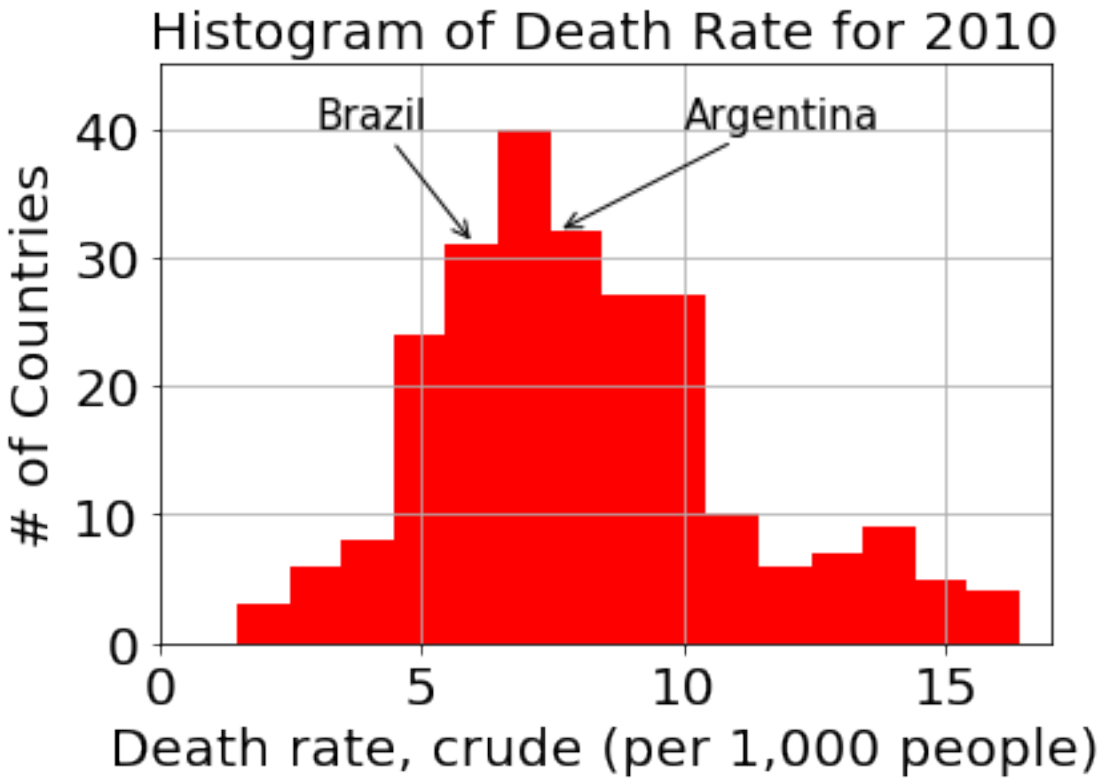
In [69]:
```
deathRate_2010[deathRate_2010['CountryCode'] == 'ARG']
```

Out[69]:

	CountryName	CountryCode	IndicatorName	IndicatorCode	Year	Value
4866686	Argentina	ARG	Death rate, crude (per 1,000 people)	SP.DYN.CDRT.IN	2010	7.64

In [70]:
```
# let's plot a histogram of the deth rate by country

# subplots returns a touple with the figure, axis attributes.
```

```
#fig, ax = plt.subplots(figsize=(15,10))
fig, ax = plt.subplots()

ax.annotate("Brazil",
            xy=(6, 31), xycoords='data', fontsize=15,
            xytext=(3, 40), textcoords='data',
            arrowprops=dict(arrowstyle="->",
                            connectionstyle="arc3"),
            )

ax.annotate("Argentina",
            xy=(7.6, 32), xycoords='data', fontsize=15,
            xytext=(10, 40), textcoords='data',
            arrowprops=dict(arrowstyle="->",
                            connectionstyle="arc3"),
            )

plt.hist(deathRate_2010['Value'], 15, density=False, facecolor='red')

plt.xlabel(deathRate_2010['IndicatorName'].iloc[0],fontsize=20)
plt.ylabel('# of Countries',fontsize=20)
plt.title('Histogram of Death Rate for 2010',fontsize=20)

plt.axis([0, 17, 0, 45])
plt.grid(True)
plt.tick_params(labelsize=20)

#plt.savefig('hist_DeathRate_2010.jpg', bbox_inches='tight')
plt.show()
```



In [71]:
```
# select Hospital beds for all countries in 2010
hist_indicator2 = 'Hospital beds'
hist_year = 2010

mask11 = data['IndicatorName'].str.contains(hist_indicator2)
mask2 = data['Year'].isin([hist_year])

# apply our mask
hospBed_2010 = data[mask11 & mask2]
hospBed_2010.head()
```

Out[71]:

	CountryName	CountryCode	IndicatorName	IndicatorCode	Year	Value
4839785	Caribbean small states	CSS	Hospital beds (per 1,000 people)	SH.MED.BEDS.ZS	2010	2.380990
4840379	Central Europe and the Baltics	CEB	Hospital beds (per 1,000 people)	SH.MED.BEDS.ZS	2010	6.497921
4841033	East Asia & Pacific (all income levels)	EAS	Hospital beds (per 1,000 people)	SH.MED.BEDS.ZS	2010	2.985314
4841702	East Asia & Pacific (developing only)	EAP				
4842503	Euro area	EMU				

```
4841702  Hospital beds (per 1,000 people)  SH.MED.BEDS.ZS  2010  2.974029
4842503  Hospital beds (per 1,000 people)  SH.MED.BEDS.ZS  2010  5.740205
```

For how many countries do we have Hospital Beds data in 2010

```
In [72]: print(len(hospBed_2010))
```

```
112
```

```
In [73]: hospBed_2010[hospBed_2010['CountryCode'] == 'BRA']
```

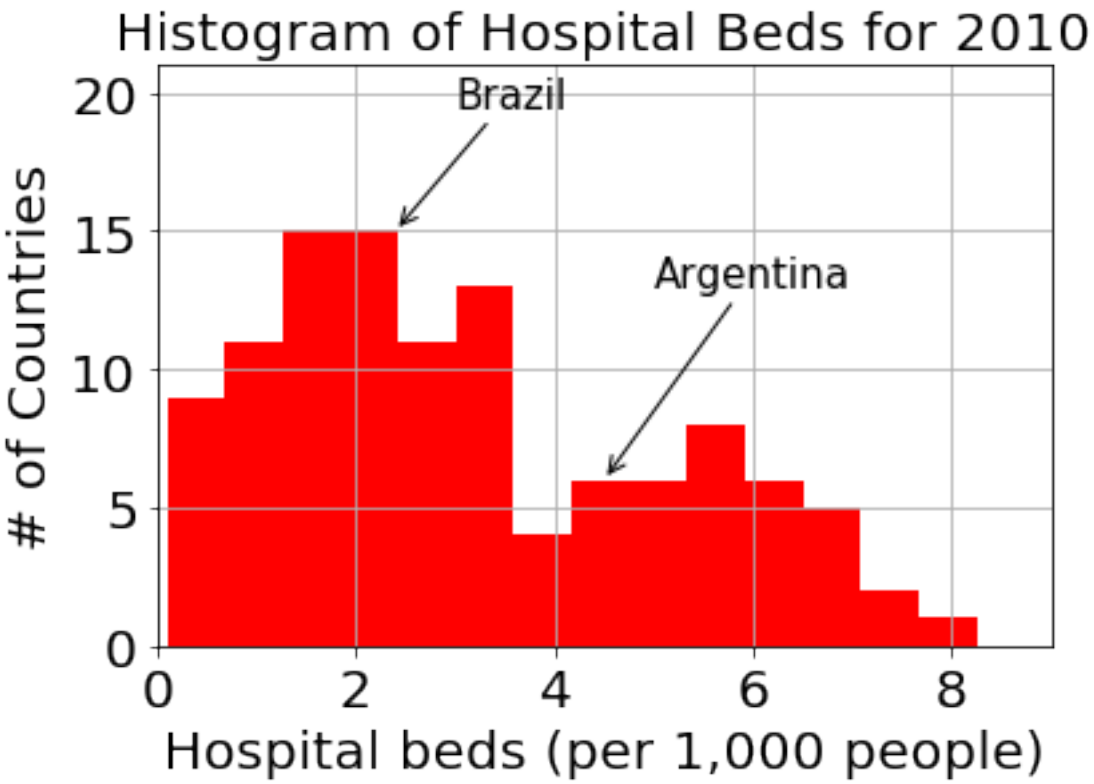
```
Out[73]:         CountryName CountryCode                     IndicatorName  \
4882748      Brazil         BRA  Hospital beds (per 1,000 people)

          IndicatorCode  Year  Value
4882748  SH.MED.BEDS.ZS  2010    2.4
```

```
In [74]: hospBed_2010[hospBed_2010['CountryCode'] == 'ARG']
```

```
Out[74]:         CountryName CountryCode                     IndicatorName  \
4866890   Argentina         ARG  Hospital beds (per 1,000 people)

          IndicatorCode  Year  Value
4866890  SH.MED.BEDS.ZS  2010    4.5
```

```
In [75]: # let's plot a histogram of the hospital beds by country

         # subplots returns a touple with the figure, axis attributes.
         #fig, ax = plt.subplots(figsize=(15,10))
         fig, ax = plt.subplots()

         ax.annotate("Brazil",
                     xy=(2.4, 15), xycoords='data', fontsize=15,
                     xytext=(3, 19.5), textcoords='data',
                     arrowprops=dict(arrowstyle="->",
                                     connectionstyle="arc3"),
                     )

         ax.annotate("Argentina",
                     xy=(4.5, 6), xycoords='data', fontsize=15,
                     xytext=(5, 13), textcoords='data',
                     arrowprops=dict(arrowstyle="->",
                                     connectionstyle="arc3"),
                     )

         plt.hist(hospBed_2010['Value'], 14, density=False, facecolor='red')

         plt.xlabel(hospBed_2010['IndicatorName'].iloc[0],fontsize=20)
```

```
plt.ylabel('# of Countries',fontsize=20)
plt.title('Histogram of Hospital Beds for 2010',fontsize=20)

plt.axis([0, 9, 0, 21])
plt.grid(True)
plt.tick_params(labelsize=20)

#plt.savefig('hist_HospBed_2010.jpg', bbox_inches='tight')
plt.show()
```



In [76]:
```
# select Maternal mortality ratio for all countries in 2010
hist_indicator3 = 'Maternal mortality ratio'
hist_year = 2010

mask3 = data['IndicatorName'].str.contains(hist_indicator3)
mask2 = data['Year'].isin([hist_year])

# apply our mask
maternalMort_2010 = data[mask3 & mask2]
maternalMort_2010.head()
```

Out[76]:

	CountryName	CountryCode	\
4839325	Arab World	ARB	

```
4839871                   Caribbean small states         CSS
4840495            Central Europe and the Baltics         CEB
4841133  East Asia & Pacific (all income levels)         EAS
4841823    East Asia & Pacific (developing only)         EAP

                                                IndicatorName IndicatorCode  \
4839325  Maternal mortality ratio (modeled estimate, pe...   SH.STA.MMRT
4839871  Maternal mortality ratio (modeled estimate, pe...   SH.STA.MMRT
4840495  Maternal mortality ratio (modeled estimate, pe...   SH.STA.MMRT
4841133  Maternal mortality ratio (modeled estimate, pe...   SH.STA.MMRT
4841823  Maternal mortality ratio (modeled estimate, pe...   SH.STA.MMRT

         Year  Value
4839325  2010  174.0
4839871  2010  104.0
4840495  2010   11.0
4841133  2010   74.0
4841823  2010   79.0
```

For how many countries do we have Maternal mortality ratio data in 2010

```
In [77]: print(len(maternalMort_2010))
```

```
265
```

```
In [78]: maternalMort_2010[maternalMort_2010['CountryCode'] == 'BRA']
```

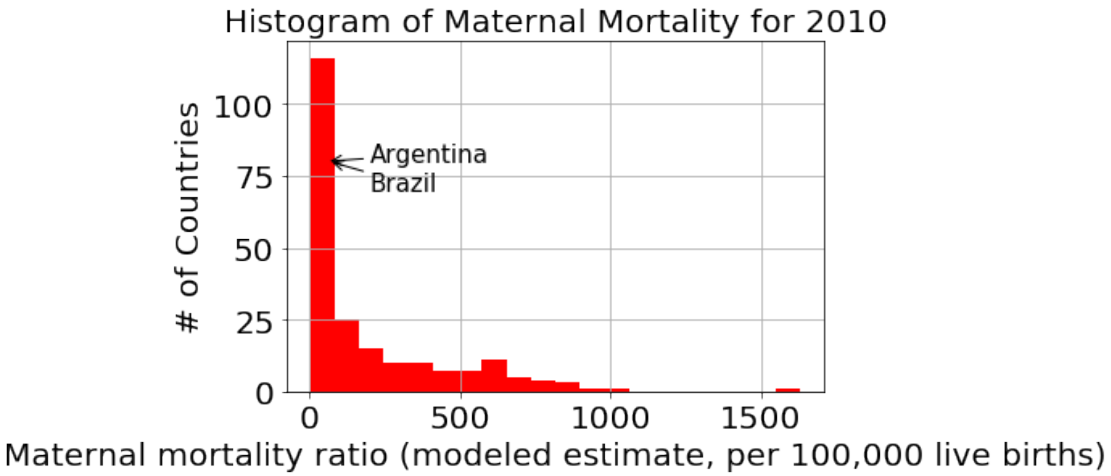
```
Out[78]:         CountryName CountryCode  \
         4882888      Brazil         BRA

                                                IndicatorName IndicatorCode  \
         4882888  Maternal mortality ratio (modeled estimate, pe...   SH.STA.MMRT

                  Year  Value
         4882888  2010   65.0
```

```
In [79]: maternalMort_2010[maternalMort_2010['CountryCode'] == 'ARG']
```

```
Out[79]:         CountryName CountryCode  \
         4867031   Argentina         ARG
         4867032   Argentina         ARG

                                                IndicatorName   IndicatorCode  \
         4867031  Maternal mortality ratio (modeled estimate, pe...     SH.STA.MMRT
         4867032  Maternal mortality ratio (national estimate, p...  SH.STA.MMRT.NE

                  Year  Value
         4867031  2010   58.0
         4867032  2010   44.0
```

In [80]:
```
maternal_Mort_2010_unique = maternalMort_2010[maternalMort_2010.IndicatorCode != 'SH.S
```

In [81]:
```
# let's plot a histogram of the hospital beds by country

# subplots returns a touple with the figure, axis attributes.
#fig, ax = plt.subplots(figsize=(15, 10))
fig, ax = plt.subplots()

ax.annotate("Brazil",
            xy=(65, 80), xycoords='data', fontsize=15,
            xytext=(200, 70), textcoords='data',
            arrowprops=dict(arrowstyle="->",
                            connectionstyle="arc3"),
            )

ax.annotate("Argentina",
            xy=(58, 80), xycoords='data', fontsize=15,
            xytext=(200, 80), textcoords='data',
            arrowprops=dict(arrowstyle="->",
                            connectionstyle="arc3"),
            )

plt.hist(maternal_Mort_2010_unique['Value'], 20, density=False, facecolor='red')

plt.xlabel(maternal_Mort_2010_unique['IndicatorName'].iloc[0], fontsize=20)
plt.ylabel('# of Countries', fontsize=20)
plt.title('Histogram of Maternal Mortality for 2010',fontsize=20)

#plt.axis([0, 1700, 0, 150])
plt.grid(True)
plt.tick_params(labelsize=20)

#plt.savefig('hist_MaternalMort_2010.jpg', bbox_inches='tight')
plt.show()
```



```
In [82]: # select Government expenditure on education for all countries in 2010
         hist_indicator4 = 'Government expenditure on education as '
         hist_year = 2010

         mask4 = data['IndicatorName'].str.contains(hist_indicator4)
         mask2 = data['Year'].isin([hist_year])

         # apply our mask
         govExp_2010 = data[mask4 & mask2]
         govExp_2010.head()
```

```
Out[82]:          CountryName CountryCode  \
         4862388  Afghanistan         AFG
         4865036      Andorra         ADO
         4865435       Angola         AGO
         4866826    Argentina         ARG
         4867740      Armenia         ARM

                                                IndicatorName      IndicatorCode  \
         4862388  Government expenditure on education as % of GD...  SE.XPD.TOTL.GD.ZS
         4865036  Government expenditure on education as % of GD...  SE.XPD.TOTL.GD.ZS
         4865435  Government expenditure on education as % of GD...  SE.XPD.TOTL.GD.ZS
         4866826  Government expenditure on education as % of GD...  SE.XPD.TOTL.GD.ZS
         4867740  Government expenditure on education as % of GD...  SE.XPD.TOTL.GD.ZS

                  Year    Value
         4862388  2010  4.51116
         4865036  2010  3.06580
         4865435  2010  3.47644
         4866826  2010  4.60777
         4867740  2010  3.24900
```

For how many countries do we have Government expenditure data in 2010

```
In [83]: print(len(govExp_2010))
```

```
120
```

```
In [84]: govExp_2010[govExp_2010['CountryCode'] == 'BRA']
```

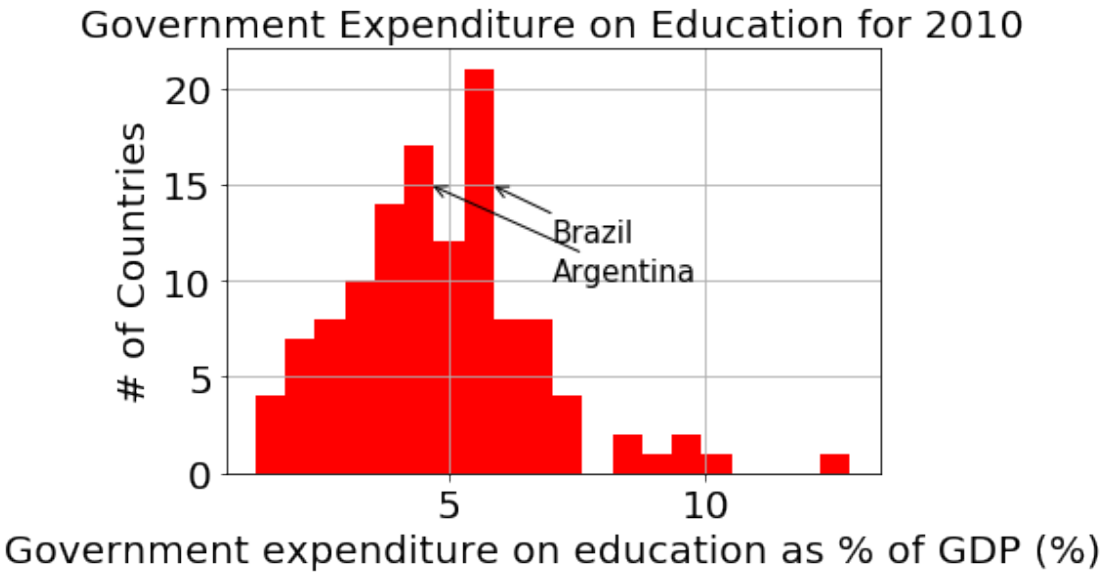
```
Out[84]:          CountryName CountryCode  \
         4882699      Brazil         BRA

                                                IndicatorName      IndicatorCode  \
         4882699  Government expenditure on education as % of GD...  SE.XPD.TOTL.GD.ZS
```

```
          Year    Value
4882699   2010  5.82225
```

In [85]:
```
govExp_2010[govExp_2010['CountryCode'] == 'ARG']
```

Out[85]:
```
         CountryName CountryCode  \
4866826   Argentina         ARG

                                                IndicatorName      IndicatorCode  \
4866826  Government expenditure on education as % of GD...  SE.XPD.TOTL.GD.ZS

          Year    Value
4866826   2010  4.60777
```

In [86]:
```
# let's plot a histogram of the hospital beds by country

# subplots returns a touple with the figure, axis attributes.
#fig, ax = plt.subplots(figsize=(15, 10))
fig, ax = plt.subplots()

ax.annotate("Brazil",
            xy=(5.8, 15), xycoords='data', fontsize=15,
            xytext=(7, 12), textcoords='data',
            arrowprops=dict(arrowstyle="->",
                            connectionstyle="arc3"),
            )

ax.annotate("Argentina",
            xy=(4.6, 15), xycoords='data', fontsize=15,
            xytext=(7, 10), textcoords='data',
            arrowprops=dict(arrowstyle="->",
                            connectionstyle="arc3"),
            )

plt.hist(govExp_2010['Value'], 20, density=False, facecolor='red')

plt.xlabel(govExp_2010['IndicatorName'].iloc[0],fontsize=20)
plt.ylabel('# of Countries',fontsize=20)
plt.title('Government Expenditure on Education for 2010',fontsize=20)

#plt.axis([0, 1700, 0, 150])
plt.grid(True)
plt.tick_params(labelsize=20)

#plt.savefig('hist_GovExp_2010.jpg', bbox_inches='tight')
plt.show()
```


Government Expenditure on Education for 2010
# of Countries
0
5
10
15
20
5
10
Brazil
Argentina
Government expenditure on education as % of GDP (%)