

## Objectifs

- \* Calculer les taux de suicide dans différents pays du Monde
- \* Sélectionner les cas extrêmes et les pays intéressants
- \* Comparer l'évolution des taux de suicide dans ces différents pays
- \* Tenter de corrélér les évolutions avec différents facteurs

Penser à utiliser le jeu de données sans valeurs aberrantes, data avec PIB, IDH, etc

In [1]:

```
import pandas as pd
import pylab as pl
import numpy as np
import matplotlib
import matplotlib.pyplot as plt
```

In [2]:

```
dfBS = pd.read_csv('./Data/Both_Sexes_10y.csv', sep=',')
dfBS = dfBS.drop('Unnamed: 0', axis = 1)
dfBS
```

Out[2]:

	Country	10-19 years	20-29 years	30-39 years	40-49 years	50-59 years	60-69 years	70-79 years	80+ years	inhabit
0	Afghanistan	3.1	10.2	9.2	6.6	5.6	5.5	11.0	42.0	3538312
1	Albania	5.0	6.5	6.1	9.1	7.8	6.0	8.3	16.3	287610
2	Algeria	1.3	4.2	5.3	4.7	4.1	4.2	5.6	9.4	4055140
3	Angola	2.6	6.6	5.4	7.0	14.8	23.8	42.1	63.5	2884248
4	Argentina	8.6	15.3	10.8	9.8	9.5	10.0	11.6	15.8	4359036
...	...	...	...	...	...	...	...	...	...	...
139	United States of America	6.3	18.0	19.2	19.3	21.7	17.9	18.4	25.1	32307134
140	Uruguay	8.1	23.0	20.6	22.0	22.1	23.6	32.7	38.9	342413
141	Uzbekistan	9.0	10.6	9.3	9.7	7.5	7.1	6.6	4.8	3184790
142	Zambia	2.8	8.3	8.9	11.8	18.7	25.8	48.9	79.7	1636350
143	Zimbabwe	4.6	11.3	13.7	19.2	29.4	41.3	81.5	205.7	1403039

144 rows × 10 columns

In [3]:

```
dfGlobal = pd.read_csv('./Data/dataGlobal.csv', sep=',')
dfGlobal.drop('Unnamed: 0', axis = 1, inplace = True)

dfGlobal = dfGlobal.sort_values(by = 'Age standardized suicide rate')
dfGlobal.head()
```

Out [3]:

	Country	HDI in 2016	HDI Rank (2017)	Age standardized suicide rate	PIB \$	Nb inhabitants	PIB/hab
12	Barbados	0.799	58	0.4	4.529050e+09	285796.0	15847.142717
4	Antigua and Barbuda	0.778	70	0.5	1.464630e+09	94527.0	15494.305472
56	Grenada	0.770	75	1.7	1.061631e+09	110261.0	9628.347487
71	Jamaica	0.732	97	2.0	1.407214e+10	2906238.0	4842.045011
77	Kuwait	0.804	56	2.2	1.094201e+11	3956873.0	27653.164401

In [4]:

```
dfGlobal.describe()
```

Out [4]:

	HDI in 2016	HDI Rank (2017)	Age standardized suicide rate	PIB \$	Nb inhabitants	PII
count	156.000000	156.000000	156.000000	1.560000e+02	1.560000e+02	156.00
mean	0.707936	94.025641	9.771154	4.414584e+11	4.276798e+07	13048.27
std	0.154896	55.222944	5.640613	1.818580e+12	1.572138e+08	18214.09
min	0.351000	1.000000	0.400000	1.783289e+08	9.452700e+04	282.14
25%	0.585750	46.750000	5.700000	1.005786e+10	2.580291e+06	1923.20
50%	0.740500	92.000000	8.900000	3.855999e+10	9.315888e+06	5147.62
75%	0.823500	143.250000	12.900000	2.100605e+11	2.799294e+07	15006.35
max	0.951000	189.000000	30.200000	1.870719e+13	1.378665e+09	101305.52

Sur les 142 pays étudiés, le taux moyen de suicide, standardisé par l'âge, en 2016, est de **10 suicides pour 100 000 habitants par an** quelque soit l'âge ou le pays

In [5]:

```
dfGlobal[dfGlobal['Country'] == 'Guyana']
#Etude d'un pays avec un des plus forts taux de suicide du monde
```

Out [5]:

	Country	HDI in 2016	HDI Rank (2017)	Age standardized suicide rate	PIB \$	Nb inhabitants	PIB/hab	
88	Guyana	0.650	185	20.0	0.504001e+10	774000.0	4540.000000	

```
60  Guyana  0.652  125  30.2  3.504024e+09  771366.0  4542.622067
```

In [6]:

```
dfGlobal[dfGlobal['Country'] == 'Greece']  
#Etude d'un pays avec un des plus faibles taux de suicide du monde
```

Out[6]:

	Country	HDI in 2016	HDI Rank (2017)	Age standardized suicide rate	PIB \$	Nb inhabitants	PIB/hab
55	Greece	0.868	31	3.8	1.952224e+11	10775971.0	18116.459622

## Etude de la France

Population totale par sexe et âge au 1er janvier 2019, France métropolitaine Mis à jour : janvier 2019  
Champ : France métropolitaine Source : Insee, estimations de population (résultats provisoires à fin 2018)

Taux de suicide associé par catégorie d'âge

In [7]:

```
dfFrance = pd.read_csv('./Data/age_France.csv', sep=',')  
dfFrance
```

Out[7]:

	Critère	10-19 years	20-29 years	30-39 years	40-49 years	50-59 years	60-69 years	7 y
0	Nb hab	7958885.0	7306158.0	8035336.0	8388238.0	8532078.0	7803284.0	53318
1	Taux suicide	2.2	8.0	13.4	21.8	23.2	20.3	

In [8]:

```
dfFranceT = dfFrance.transpose()  
dfFranceT = dfFranceT.drop('Critère', axis = 0)  
dfFranceT
```

Out[8]:

	0	1
10-19 years	7.95888e+06	2.2
20-29 years	7.30616e+06	8
30-39 years	8.03534e+06	13.4
40-49 years	8.38824e+06	21.8
50-59 years	8.53208e+06	23.2
60-69 years	7.80328e+06	20.3
70-79 years	5.3318e+06	27.2

80+ years 3.96985e+06 73.4

In [9]:

```
dfFranceT.iloc[:,1] = pd.to_numeric(dfFranceT.iloc[:,1])
dfFranceT['Taux standardisé par âge pour 100000 habitants'] = dfFranceT[1]
*100000 / dfFranceT[0]
#pour 100 000 habitants

dfFranceT['Age'] = [15,25,35,45,55,65,75,85]

dfFranceT.rename(columns={0: 'Nb inhabitant', 1:'Taux de suicide'}, inplace=True)
```

In [10]:

dfFranceT

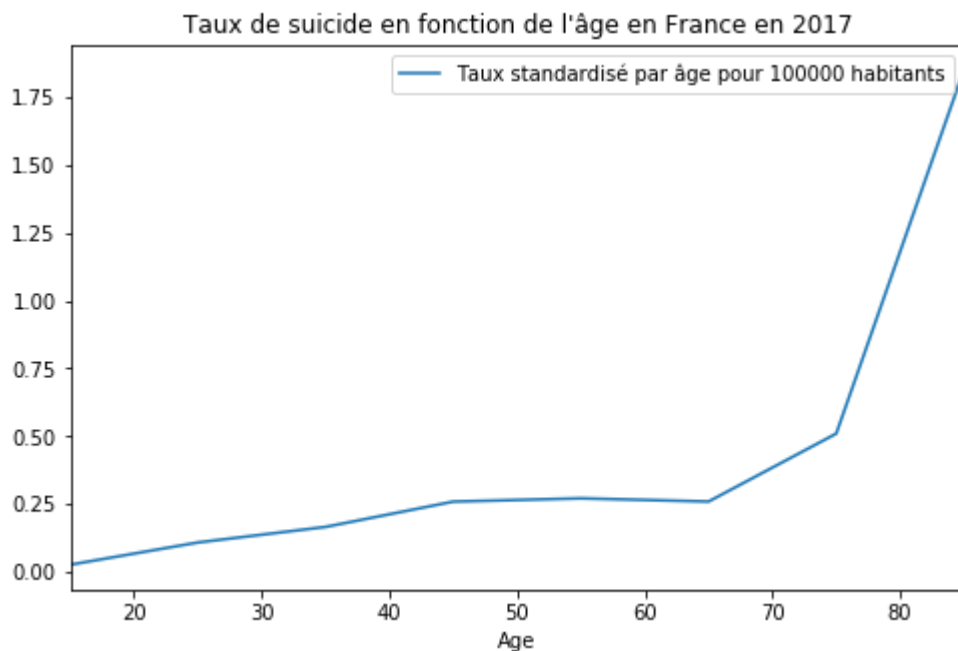
Out[10]:

	Nb inhabitant	Taux de suicide	Taux standardisé par âge pour 100000 habitants	Age
10-19 years	7.95888e+06	2.2	0.0276421	15
20-29 years	7.30616e+06	8.0	0.109497	25
30-39 years	8.03534e+06	13.4	0.166763	35
40-49 years	8.38824e+06	21.8	0.259888	45
50-59 years	8.53208e+06	23.2	0.271915	55
60-69 years	7.80328e+06	20.3	0.260147	65
70-79 years	5.3318e+06	27.2	0.510147	75
80+ years	3.96985e+06	73.4	1.84894	85

In [11]:

```
#création d'une fenêtre Matplotlib
fig, ax = plt.subplots(1, figsize=(8,5));
dfFranceT.plot(x = 'Age', y = 'Taux standardisé par âge pour 100000 habitant
s', ax = ax);
ax.set_title("Taux de suicide en fonction de l'âge en France en 2017");

plt.savefig('NB4_0.png');
```



In [12]:

```
dfFranceT.to_csv('./Data/dataFrance.csv')
```

On constate que le taux de suicide explose pour les plus de 80ans.

On peut mettre en cause la mauvaise prise en charge des personnes âgées, leurs conditions de vie, la solitude et le sentiment d'exclusion de la société

Dans d'autres cultures, principalement au Maghreb (Afrique du Nord), la personne âgée est respectée, elle fait figure de sagesse et est le témoin d'une génération entière. Accueillies dans leur famille lorsqu'elle ne sont plus autonomes, les seniors conservent une place à part entière dans la société, quelque soit leur âge

In [13]:

```
dfBS = dfBS.sort_values(by = "80+ years")
dfBS.head(10)
```

Out[13]:

	Country	10-19 years	20-29 years	30-39 years	40-49 years	50-59 years	60-69 years	70-79 years	80+ years	Population inhabitant
70	Kuwait	0.8	4.3	3.4	2.1	2.3	2.6	4.9	0.0	3956873
83	Mauritius	3.7	12.9	11.8	11.0	7.4	6.4	5.5	4.0	1263473
141	Uzbekistan	9.0	10.6	9.3	9.7	7.5	7.1	6.6	4.8	31847900

65	Jamaica	1.2	2.8	2.5	2.7	2.9	2.8	3.7	5.3	2906236
138	United Arab Emirates	1.8	4.0	2.8	2.8	3.2	4.4	5.6	6.7	9360980
34	Cyprus	1.6	6.2	6.1	6.2	6.3	8.9	7.4	6.8	1170187
51	Greece	0.7	4.0	4.3	6.4	7.5	7.1	6.4	7.2	10775971
52	Guatemala	3.1	4.3	3.2	3.5	3.1	4.1	2.8	7.4	16583060
99	Pakistan	3.1	5.5	3.8	2.9	2.7	2.9	4.0	7.5	203627284
8	Azerbaijan	1.4	2.2	3.2	3.5	4.1	4.5	6.6	7.7	9757812

Nous comparons avec une carte sur le Notebook 4 bis afin de confirmer ou infirmer cette hypothèse. Pour autant, ces données ne sont pas réellement représentatives car elles sont pondérées par le nombre total d'habitant et non pas par le nombre de personnes de + de 80 ans

## Etude de trois "gros" pays

### Chine, Etats-Unis et Inde

In [14]:

```
dfBS = dfBS.sort_values(by = "Nb inhabitant", ascending= False )
dfBS.head(3)
```

Out[14]:

	Country	10-19 years	20-29 years	30-39 years	40-49 years	50-59 years	60-69 years	70-79 years	80+ years	Nb inhabitan
28	China	1.9	5.0	6.0	6.3	10.5	23.4	44.0	61.3	1.378665e+09
59	India	11.2	30.3	22.9	17.7	16.8	17.5	21.9	24.5	1.324510e+09
139	United States of America	6.3	18.0	19.2	19.3	21.7	17.9	18.4	25.1	3.230713e+08

Pourquoi étudier ces 3 pays ? Ces pays sont les plus peuplés du monde, nous espérons donc que la forte population nous permette d'avoir des valeurs plus fiables liées à une taille d'échantillon plus élevée.

In [15]:

```
dfGlobal = dfGlobal.sort_values(by = "Nb inhabitants", ascending= False )
dfGlobal.head(3)
```

Out[15]:

	Country	HDI in 2016	HDI Rank (2017)	Age standardized suicide rate	PIB \$	Nb inhabitants	PIB/ha
32	China	0.748	86	8.0	1.113795e+13	1.378665e+09	8078.79047
59	India	0.600	180	10.5	2.800000e+12	1.324510e+09	1700.00000

65	India	0.636	130	16.5	2.290432e+12	1.324510e+09	1729.26802
150	United States	0.922	13	13.7	1.870719e+13	3.230713e+08	57904.20196

In [16]:

```
dfMean = dfBS.describe()
dfMean = dfMean.drop('Nb_inhabitant', axis = 1).transpose()
dfMean['Age'] = [15, 25, 35, 45, 55, 65, 75, 85]
dfMean
```

Out[16]:

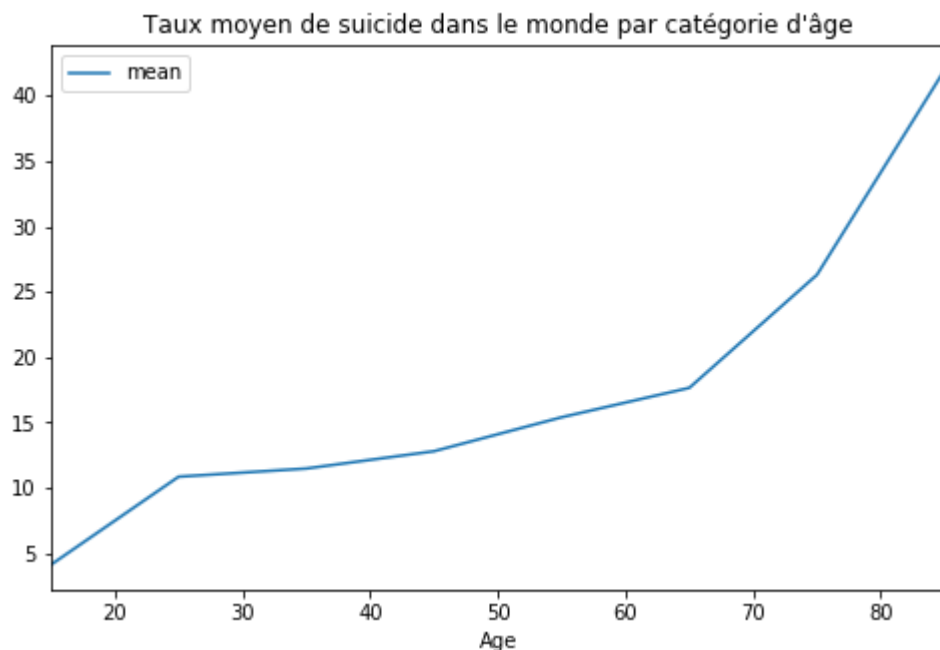
	count	mean	std	min	25%	50%	75%	max	Age
10-19 years	144.0	4.134028	2.852680	0.7	2.175	3.40	5.300	18.3	15
20-29 years	144.0	10.865278	6.958753	2.2	5.650	8.90	14.075	38.0	25
30-39 years	144.0	11.488889	7.999983	2.5	6.075	9.15	14.225	49.7	35
40-49 years	144.0	12.811806	8.652357	2.1	6.675	10.50	16.700	54.8	45
50-59 years	144.0	15.409722	9.859093	2.3	7.575	12.75	21.700	48.5	55
60-69 years	144.0	17.658333	11.522435	2.6	8.275	16.35	23.725	55.4	65
70-79 years	144.0	26.306250	21.108366	2.8	9.800	19.15	36.175	99.1	75
80+ years	144.0	41.976389	35.705777	0.0	15.775	32.60	54.100	205.7	85

In [17]:

```
fig, ax = plt.subplots(1, figsize=(8,5));
dfMean.plot(x = 'Age', y = 'mean', ax = ax);
ax.set_title("Taux moyen de suicide dans le monde par catégorie d'âge");

plt.savefig('NB4_1.png');

#résultat non exploitable car pas de pondération par catégorie d'âge
```



Afin de comparer ces trois pays, nous allons comparer l'évolution des taux de suicide en fonction des

années, comparativement à la moyenne mondiale.

Par une étude du PIB, de l'IDH et d'autres critères, nous essayerons d'interpréter ces résultats.

In [18]:

```
df = pd.read_csv('./Data/AS_2000-2016.csv', sep=',')
df
```

Out[18]:

	Country	Sex	2016	2015	2010	2000
0	Afghanistan	Both sexes	6.4	6.6	7.4	8.1
1	Afghanistan	Male	10.6	10.9	12.5	14.3
2	Afghanistan	Female	2.1	2.1	2.1	1.7
3	Albania	Both sexes	5.6	5.3	7.7	5.8
4	Albania	Male	7.0	6.7	9.5	8.2
...	...	...	...	...	...	...
544	Zambia	Male	17.5	17.4	17.9	21.9
545	Zambia	Female	6.2	6.1	6.2	7.5
546	Zimbabwe	Both sexes	19.1	18.9	20.6	21.7
547	Zimbabwe	Male	29.1	28.7	32.3	35.5
548	Zimbabwe	Female	11.1	11.1	11.1	9.8

549 rows × 6 columns

In [19]:

```
#passer les valeurs de cdc à numérique
for i in range(2,6) :
    df.iloc[:,i] = pd.to_numeric(df.iloc[:,i])
df.head(20)

#création de 3 df : Both Sexes, Male et Female
dfAS = df[df['Sex'] == 'Both sexes'].drop('Sex', axis = 1)
dfAS.index = range(0, dfAS.shape[0])
dfMale = df[df['Sex'] == 'Male'].drop('Sex', axis = 1)
dfMale.index = range(0, dfMale.shape[0])
dfFemale = df[df['Sex'] == 'Female'].drop('Sex', axis = 1)
dfFemale.index = range(0, dfFemale.shape[0])

dfAS.head(5)
```

Out[19]:

	Country	2016	2015	2010	2000
0	Afghanistan	6.4	6.6	7.4	8.1
1	Albania	5.6	5.3	7.7	5.8
2	Algeria	3.3	3.4	3.5	4.7
3	Angola	8.9	9.3	10.4	13.9



```
4 Antigua and Barbuda    0.5    0.8    0.2    2.1
```

In [20]:

```
dfIh = pd.read_csv('./Data/inhabitants.csv', sep=',')
dfIh.iloc[:,1] = pd.to_numeric(dfIh.iloc[:,1])
dfIh.rename(columns={'2016': 'Nb inhabitant'}, inplace=True)

dfAS = dfAS.merge(dfIh, left_on = 'Country', right_on = 'Country')
dfAS = dfAS.sort_values(by = 'Nb inhabitant', ascending = False)
dfAS
```

Out[20]:

	Country	2016	2015	2010	2000	Nb inhabitant
32	China	8.0	8.1	10.0	14.1	1.378665e+09
66	India	16.5	16.7	17.1	18.7	1.324510e+09
153	United States of America	13.7	13.3	11.8	10.1	3.230713e+08
67	Indonesia	3.7	3.7	4.0	4.3	2.615542e+08
20	Brazil	6.1	6.2	5.7	5.3	2.061631e+08
...	...	...	...	...	...	...
57	Grenada	1.7	1.1	0.9	2.7	1.102610e+05
145	Tonga	4.0	4.0	4.3	4.5	1.011330e+05
126	Seychelles	8.3	8.5	9.0	9.4	9.467700e+04
4	Antigua and Barbuda	0.5	0.8	0.2	2.1	9.452700e+04
45	Eritrea	13.8	14.1	15.0	18.9	NaN

159 rows × 6 columns

In [21]:

```
#On s'intéresse au tableau des taux de suicide "Age Standardized" Réduit à
nos trois pays
#Estimation des taux de suicide, age-standardized Estimations par pays et
par sexe, pour 100 000 hab

dfASR = dfAS.head(4)
dfASR = dfASR.drop('Nb inhabitant', axis = 1)
dfASR = dfASR.transpose()
dfASR = dfASR.drop('Country', axis = 0)
dfASR = dfASR.rename(columns={32: 'China', 66: 'India', 153: 'United States
of America', 67: 'Indonesia'})
dfASR['Year']=[2016,2015,2010,2000]
dfASR
```

Out[21]:

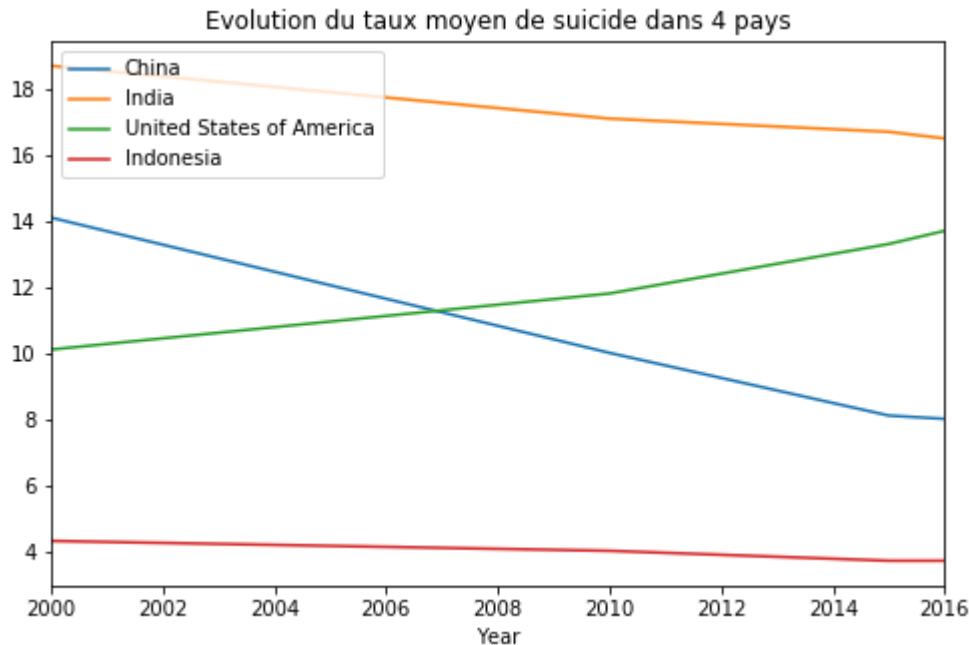
	China	India	United States of America	Indonesia	Year
2016	8	16.5	13.7	3.7	2016
2015	8.1	16.7	13.3	3.7	2015
2010	10	17.1	11.8	4	2010

2000    14.1    18.7                    10.1                    4.3    2000

In [22]:

```
fig, ax = plt.subplots(1, figsize=(8,5));
dfASR.plot(x = 'Year', ax = ax);
ax.set_title("Evolution du taux moyen de suicide dans 4 pays");

plt.savefig('NB4_2.png');
```



On remarque que, si l'Inde est le pays avec le plus fort taux de suicide, il est sur la voie d'une amélioration avec une baisse constante de ce taux depuis 2000. A l'inverse, on constate que ce taux est en forte hausse aux Etats-Unis. Contrairement à ce que l'on pourrait croire, un pays développé n'implique pas systématiquement un taux de suicide faible.

## Etude de l'évolution des taux de suicide

In [23]:

```
dfAS
```

Out [23]:

	Country	2016	2015	2010	2000	Nb inhabitant
32	China	8.0	8.1	10.0	14.1	1.378665e+09
66	India	16.5	16.7	17.1	18.7	1.324510e+09
153	United States of America	13.7	13.3	11.8	10.1	3.230713e+08
67	Indonesia	3.7	3.7	4.0	4.3	2.615542e+08
20	Brazil	6.1	6.2	5.7	5.3	2.061631e+08
...	...	...	...	...	...	...
57	Grenada	1.7	1.1	0.9	2.7	1.102610e+05
145	Tonga	4.0	4.0	4.3	4.5	1.011330e+05

126	Seychelles	8.3	8.5	9.0	9.4	9.467700e+04
4	Antigua and Barbuda	0.5	0.8	0.2	2.1	9.452700e+04
45	Eritrea	13.8	14.1	15.0	18.9	NaN

159 rows × 6 columns

In [24]:

```
dfAS['Evolution 2000-2016'] = dfAS['2016'] - dfAS['2000']
dfAS = dfAS.sort_values(by = 'Evolution 2000-2016')
dfA1 = dfAS.head(3)
dfA1

#sélection des trois pays avec la plus forte baisse du taux de suicide
```

Out[24]:

	Country	2016	2015	2010	2000	Nb inhabitant	Evolution 2000- 2016
119	Russian Federation	26.5	27.5	34.5	47.5	144342396.0	-21.0
84	Lithuania	25.7	28.8	31.3	46.3	2868231.0	-20.6
13	Belarus	21.4	21.2	31.9	38.7	9501534.0	-17.3

In [25]:

```
dfAS = dfAS.sort_values(by = 'Evolution 2000-2016', ascending = False)
dfA2 = dfAS.head(3)
dfA2

#sélection des trois pays avec la plus forte hausse du taux de suicide
```

Out[25]:

	Country	2016	2015	2010	2000	Nb inhabitant	Evolution 2000- 2016
153	United States of America	13.7	13.3	11.8	10.1	323071342.0	3.6
61	Guyana	30.2	30.3	29.3	26.7	771366.0	3.5
116	Portugal	8.6	8.6	8.8	5.3	10325452.0	3.3

In [26]:

```
dfEvo = pd.concat([dfA1, dfA2])
dfEvo = dfEvo.drop('Nb inhabitant', axis = 1)
dfEvo = dfEvo.transpose()
dfEvo = dfEvo.drop('Country', axis = 0).drop('Evolution 2000-2016', axis = 0)
dfEvo['Year']=[2016,2015,2010,2000]
dfEvo = dfEvo.rename(columns={119: 'Russie', 84: 'Lithuanie', 13: 'Belarus',
, 153: 'Etats-Unis', 61: 'Guyana', 116: 'Portugal'})
dfEvo
```

Out[26]:

	Russie	Lithuanie	Belarus	Etats-Unis	Guyana	Portugal	Year
2016	26.5	25.7	21.4	13.7	30.2	8.6	2016
2015	27.5	28.8	21.2	13.3	30.3	8.6	2015
2010	34.5	31.3	31.9	11.8	29.3	8.8	2010
2000	47.5	46.3	38.7	10.1	26.7	5.3	2000

In [27]:

```
fig, ax = plt.subplots(1, figsize=(8,5));
dfEvo.plot(x = 'Year', ax = ax);
ax.set_title("Evolution du taux moyen de suicide dans 6 pays représentatifs");
plt.savefig('NB4_3.png');
```

