US_Crime_Rates*1960*2014数据 - 函数

导入数据

```
1  import pandas as pd
```

```
1  crime = pd.read_csv('data/US_Crime_Rates_1960_2014.csv')
2  crime.info()
```

```
1   <class 'pandas.core.frame.DataFrame'>
2   RangeIndex: 55 entries, 0 to 54
3   Data columns (total 12 columns):
4   Year                55 non-null int64
5   Population          55 non-null int64
6   Total               55 non-null int64
7   Violent             55 non-null int64
8   Property            55 non-null int64
9   Murder              55 non-null int64
10  Forcible_Rape       55 non-null int64
11  Robbery             55 non-null int64
12  Aggravated_assault  55 non-null int64
13  Burglary            55 non-null int64
14  Larceny_Theft       55 non-null int64
15  Vehicle_Theft       55 non-null int64
16  dtypes: int64(12)
17  memory usage: 5.2 KB
```

将Year的数据类型转换为 datetime64 显示格式为年

```
1   crime.Year = pd.to_datetime(crime.Year, format='%Y')
2   crime.Year.head(3)
```

```
1   0   1960-01-01
2   1   1961-01-01
3   2   1962-01-01
4   Name: Year, dtype: datetime64[ns]
```

将列Year设置为数据框的索引

```
1  crime = crime.set_index('Year', drop=True) # drop 是否删除该
   列，默认为True
2  crime.head(3)
```

```
<tr style="text-align: right;">
  <th></th>
  <th>Population</th>
  <th>Total</th>
  <th>Violent</th>
  <th>Property</th>
  <th>Murder</th>
  <th>Forcible_Rape</th>
  <th>Robbery</th>
  <th>Aggravated_assault</th>
  <th>Burglary</th>
  <th>Larceny_Theft</th>
  <th>Vehicle_Theft</th>
</tr>
<tr>
  <th>Year</th>
  <th></th>
  <th></th>
  <th></th>
  <th></th>
  <th></th>
  <th></th>
  <th></th>
  <th></th>
  <th></th>
  <th></th>
  <th></th>
</tr>
```

```
<tr>
  <th>1960-01-01</th>
  <td>179323175</td>
  <td>3384200</td>
  <td>288460</td>
  <td>3095700</td>
  <td>9110</td>
  <td>17190</td>
  <td>107840</td>
  <td>154320</td>
  <td>912100</td>
  <td>1855400</td>
  <td>328200</td>
</tr>
<tr>
  <th>1961-01-01</th>
  <td>182992000</td>
  <td>3488000</td>
  <td>289390</td>
  <td>3198600</td>
  <td>8740</td>
  <td>17220</td>
  <td>106670</td>
  <td>156760</td>
  <td>949600</td>
  <td>1913000</td>
  <td>336000</td>
</tr>
<tr>
  <th>1962-01-01</th>
  <td>185771000</td>
  <td>3752200</td>
  <td>301510</td>
  <td>3450700</td>
```

```
35    <td>8530</td>
36    <td>17550</td>
37    <td>110860</td>
38    <td>164570</td>
39    <td>994300</td>
40    <td>2089600</td>
41    <td>366800</td>
42  </tr>
```

删除名为Total的列

```
1  del crime['Total']
2  crime.head(3)
```

```
<tr style="text-align: right;">
  <th></th>
  <th>Population</th>
  <th>Violent</th>
  <th>Property</th>
  <th>Murder</th>
  <th>Forcible_Rape</th>
  <th>Robbery</th>
  <th>Aggravated_assault</th>
  <th>Burglary</th>
  <th>Larceny_Theft</th>
  <th>Vehicle_Theft</th>
</tr>
<tr>
  <th>Year</th>
  <th></th>
  <th></th>
  <th></th>
  <th></th>
  <th></th>
  <th></th>
  <th></th>
  <th></th>
  <th></th>
  <th></th>
</tr>
```

```
<tr>
  <th>1960-01-01</th>
  <td>179323175</td>
  <td>288460</td>
  <td>3095700</td>
  <td>9110</td>
  <td>17190</td>
  <td>107840</td>
  <td>154320</td>
  <td>912100</td>
  <td>1855400</td>
  <td>328200</td>
</tr>
<tr>
  <th>1961-01-01</th>
  <td>182992000</td>
  <td>289390</td>
  <td>3198600</td>
  <td>8740</td>
  <td>17220</td>
  <td>106670</td>
  <td>156760</td>
  <td>949600</td>
  <td>1913000</td>
  <td>336000</td>
</tr>
<tr>
  <th>1962-01-01</th>
  <td>185771000</td>
  <td>301510</td>
  <td>3450700</td>
  <td>8530</td>
  <td>17550</td>
  <td>110860</td>
```

```
35    <td>164570</td>
36    <td>994300</td>
37    <td>2089600</td>
38    <td>366800</td>
39  </tr>
```

按照Year对数据框进行分组并求和

```python
# 关于 .resample 的介绍
# https://pandas.pydata.org/pandas-
docs/stable/generated/pandas.DataFrame.resample.html
#
https://blog.csdn.net/wangshuang1631/article/details/523149
44
# 更多关于 Offset Aliases的介绍
# (http://pandas.pydata.org/pandas-
docs/stable/timeseries.html#offset-aliases)

# AS = YS = year start
# 10AS = 以10年为周期执行重采样，获取每个10年的犯罪率
# 10年的重采样求和就是做10年数据的累加
crimes = crime.resample('10AS').sum()

# 10年的犯罪数可以累加，但是人口累加没有意义
# 用resample去得到"Population"列的最大值
population = crime['Population'].resample('10AS').max()

# 更新 "Population"
crimes['Population'] = population

crimes
```

```
<tr style="text-align: right;">
  <th></th>
  <th>Population</th>
  <th>Violent</th>
  <th>Property</th>
  <th>Murder</th>
  <th>Forcible_Rape</th>
  <th>Robbery</th>
  <th>Aggravated_assault</th>
  <th>Burglary</th>
  <th>Larceny_Theft</th>
  <th>Vehicle_Theft</th>
</tr>
<tr>
  <th>Year</th>
  <th></th>
  <th></th>
  <th></th>
  <th></th>
  <th></th>
  <th></th>
  <th></th>
  <th></th>
  <th></th>
  <th></th>
</tr>
```

```
<tr>
  <th>1960-01-01</th>
  <td>201385000.0</td>
  <td>4134930</td>
  <td>45160900</td>
  <td>106180</td>
  <td>236720</td>
  <td>1633510</td>
  <td>2158520</td>
  <td>13321100</td>
  <td>26547700</td>
  <td>5292100</td>
</tr>
<tr>
  <th>1970-01-01</th>
  <td>220099000.0</td>
  <td>9607930</td>
  <td>91383800</td>
  <td>192230</td>
  <td>554570</td>
  <td>4159020</td>
  <td>4702120</td>
  <td>28486000</td>
  <td>53157800</td>
  <td>9739900</td>
</tr>
<tr>
  <th>1980-01-01</th>
  <td>248239000.0</td>
  <td>14074328</td>
  <td>117048900</td>
  <td>206439</td>
  <td>865639</td>
  <td>5383109</td>
```

```
35    <td>7619130</td>
36    <td>33073494</td>
37    <td>72040253</td>
38    <td>11935411</td>
39  </tr>
40  <tr>
41    <th>1990-01-01</th>
42    <td>272690813.0</td>
43    <td>17527048</td>
44    <td>119053499</td>
45    <td>211664</td>
46    <td>998827</td>
47    <td>5748930</td>
48    <td>10568963</td>
49    <td>26750015</td>
50    <td>77679366</td>
51    <td>14624418</td>
52  </tr>
53  <tr>
54    <th>2000-01-01</th>
55    <td>307006550.0</td>
56    <td>13968056</td>
57    <td>100944369</td>
58    <td>163068</td>
59    <td>922499</td>
60    <td>4230366</td>
61    <td>8652124</td>
62    <td>21565176</td>
63    <td>67970291</td>
64    <td>11412834</td>
65  </tr>
66  <tr>
67    <th>2010-01-01</th>

68    <td>318857056.0</td>
```

```
69      <td>6072017</td>
70      <td>44095950</td>
71      <td>72867</td>
72      <td>421059</td>
73      <td>1749809</td>
74      <td>3764142</td>
75      <td>10125170</td>
76      <td>30401698</td>
77      <td>3569080</td>
78    </tr>
79    <tr>
80      <th>2020-01-01</th>
81      <td>NaN</td>
82      <td>0</td>
83      <td>0</td>
84      <td>0</td>
85      <td>0</td>
86      <td>0</td>
87      <td>0</td>
88      <td>0</td>
89      <td>0</td>
90      <td>0</td>
91    </tr>
```

何时是美国历史上生存最危险的年代?

```python
crime = pd.read_csv('data/US_Crime_Rates_1960_2014.csv')
crime = crime.set_index('Year', drop=True)
crime.idxmax(0)
```

```
 1   Population            2014
 2   Total                 1991
 3   Violent               1992
 4   Property              1991
 5   Murder                1991
 6   Forcible_Rape         1992
 7   Robbery               1991
 8   Aggravated_assault    1993
 9   Burglary              1980
10   Larceny_Theft         1991
11   Vehicle_Theft         1991
12   dtype: int64
```