Program 2a - Showcase the percentage of growth in population between two items (USA and China)

In [3]:
```
import pandas as pd # We’ll be using Pandas library to work with the dataset
from matplotlib import pyplot as plt
```

In [4]:
```
data=pd.read_csv('countries.csv') #read the file as data
```

In [5]:
```
data # displays the data set
```

Out[5]:

	country	year	population
0	Afghanistan	1952	8425333
1	Afghanistan	1957	9240934
2	Afghanistan	1962	10267083
3	Afghanistan	1967	11537966
4	Afghanistan	1972	13079460
...	...	...	...
1699	Zimbabwe	1987	9216418
1700	Zimbabwe	1992	10704340
1701	Zimbabwe	1997	11404948
1702	Zimbabwe	2002	11926563
1703	Zimbabwe	2007	12311143

1704 rows × 3 columns

In [6]:
```
type(data) #type of data
```

Out[6]:
```
pandas.core.frame.DataFrame
```

In [10]:
```
data.tail() # the tail command to see the last 5 items in the csv file
```

Out[10]:

	country	year	population
1699	Zimbabwe	1987	9216418
1700	Zimbabwe	1992	10704340
1701	Zimbabwe	1997	11404948
1702	Zimbabwe	2002	11926563
1703	Zimbabwe	2007	12311143

In [8]:
```
data.info() # we’ll be able to see all of the available columns in the dataset alor
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1704 entries, 0 to 1703
Data columns (total 3 columns):
 #   Column      Non-Null Count  Dtype 
---  ------      --------------  ----- 
 0   country     1704 non-null   object
 1   year        1704 non-null   int64 
 2   population  1704 non-null   int64 
dtypes: int64(2), object(1)
memory usage: 40.1+ KB
```

In [9]:
```
data.describe() # the index results include the count, mean, std, minimum 25%, 50%,
```

Out[9]:

	year	population
count	1704.00000	1.704000e+03
mean	1979.50000	2.960121e+07
std	17.26533	1.061579e+08
min	1952.00000	6.001100e+04
25%	1965.75000	2.793664e+06
50%	1979.50000	7.023596e+06
75%	1993.25000	1.958522e+07
max	2007.00000	1.318683e+09

In [17]:
```
#compare the population of US and China

#isolate the data of US and China

data.country == 'United States'#showcase when and where US as True and fast in list
```

Out[17]:
```
0       False
1       False
2       False
3       False
4       False
        ...  
1699    False
1700    False
1701    False
1702    False
1703    False
Name: country, Length: 1704, dtype: bool
```

In [ ]:
```

```

In [15]:
```
us = data[data.country == 'United States'] #segregating US data
```

In [16]:
```
us
```

Out[16]:

	country	year	population
1608	United States	1952	157553000
1609	United States	1957	171984000
1610	United States	1962	186538000
1611	United States	1967	198712000
1612	United States	1972	209896000
1613	United States	1977	220239000
1614	United States	1982	232187835
1615	United States	1987	242803533
1616	United States	1992	256894189
1617	United States	1997	272911760
1618	United States	2002	287675526
1619	United States	2007	301139947

In [18]:

```
china = data[data.country == 'China'] #segregating China's data
```

In [19]:

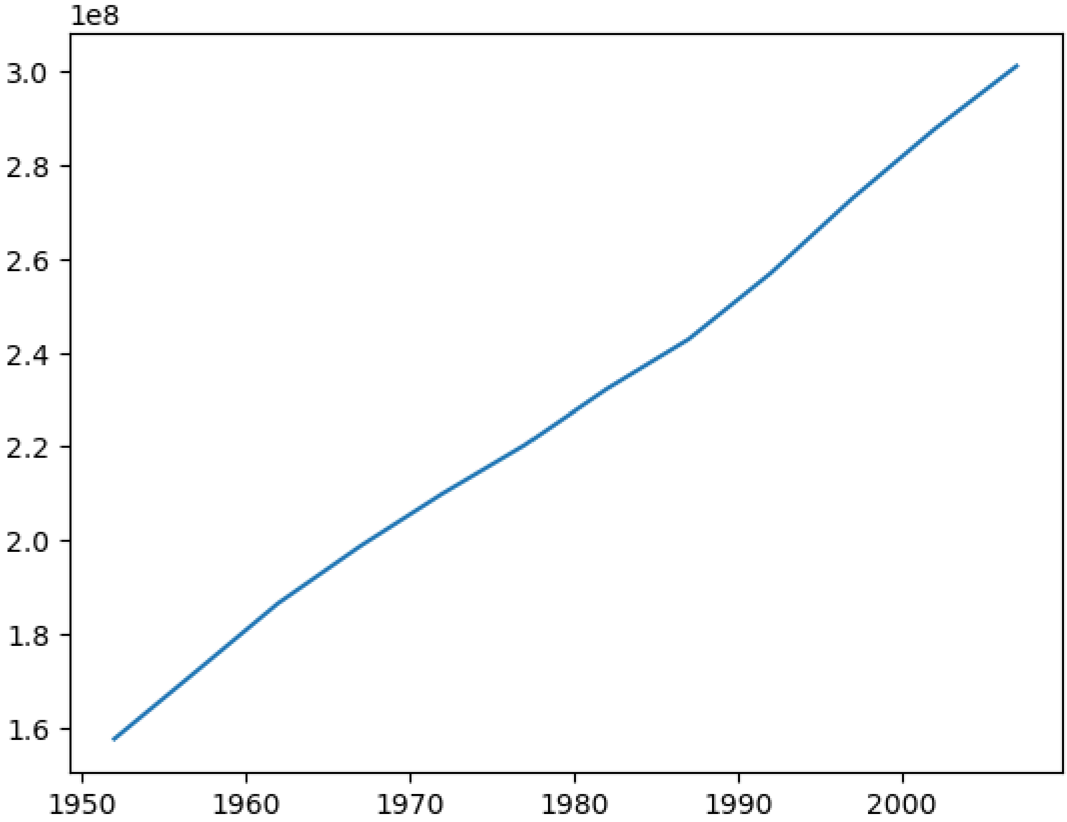
```
china
```

Out[19]:

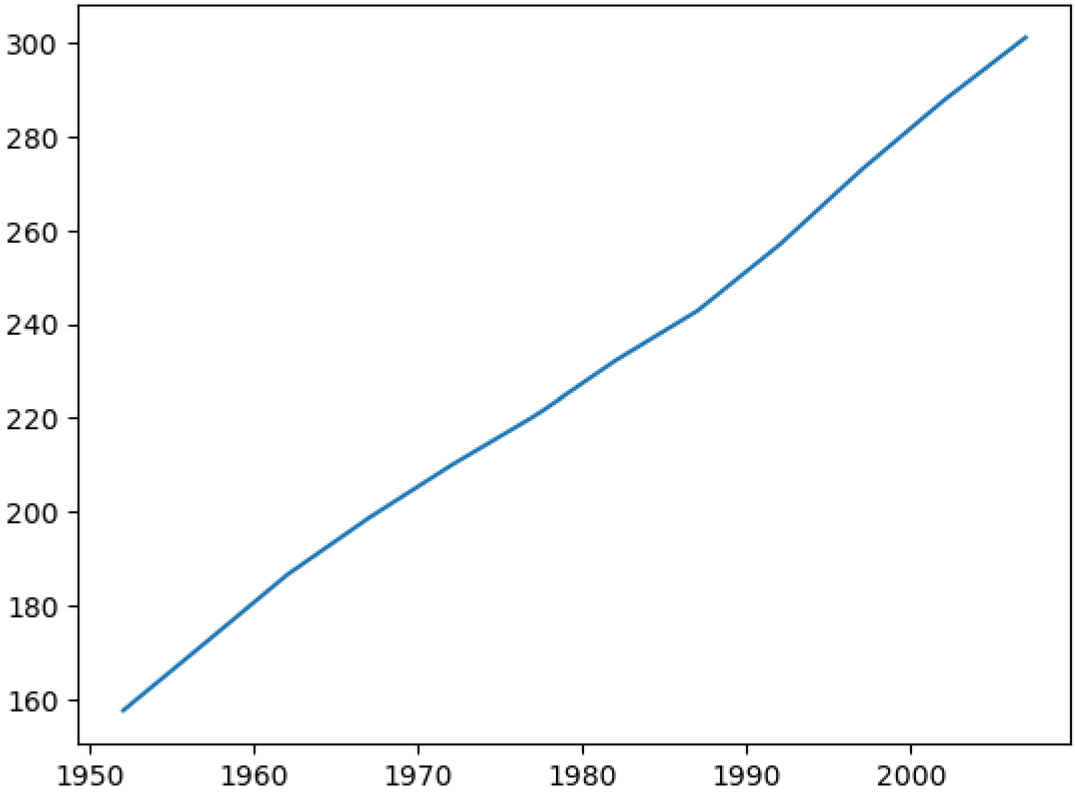
	country	year	population
288	China	1952	556263527
289	China	1957	637408000
290	China	1962	665770000
291	China	1967	754550000
292	China	1972	862030000
293	China	1977	943455000
294	China	1982	1000281000
295	China	1987	1084035000
296	China	1992	1164970000
297	China	1997	1230075000
298	China	2002	1280400000
299	China	2007	1318683096

In [20]:

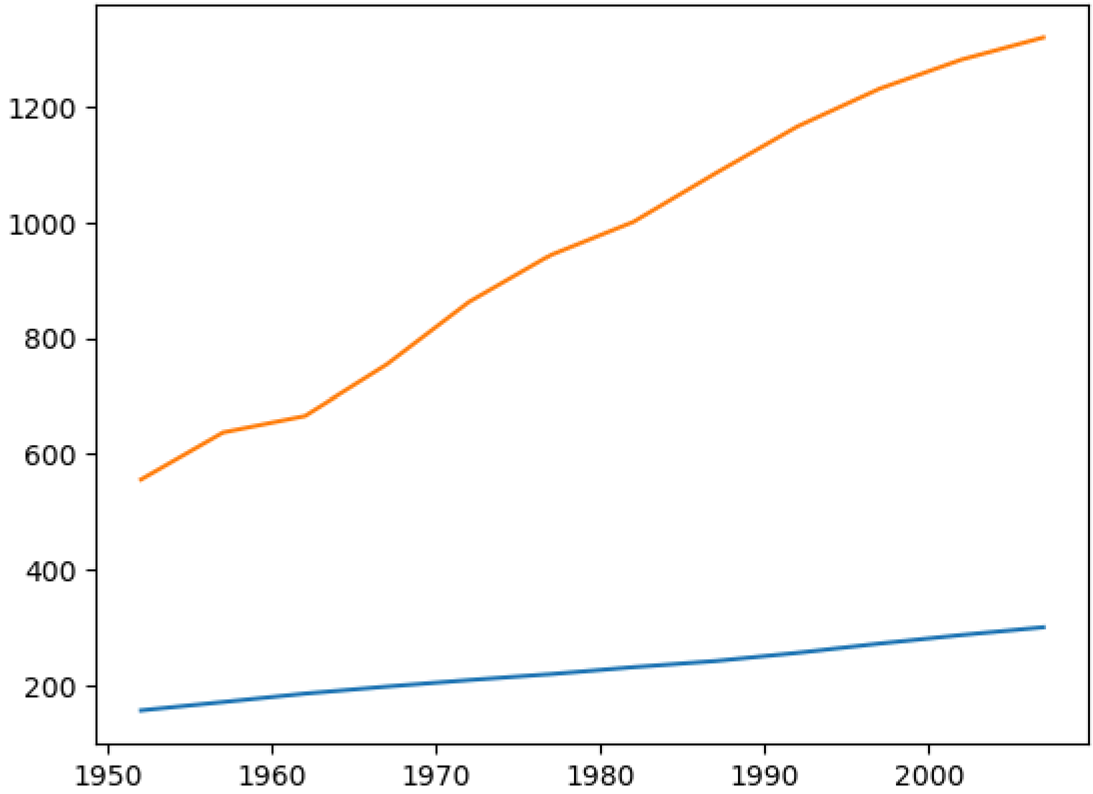
```
plt.plot(us.year, us.population)
plt.show ()
```



In [21]:
```
plt.plot(us.year, us.population / 10**6) # divide the population by 1 million
plt.show ()
```

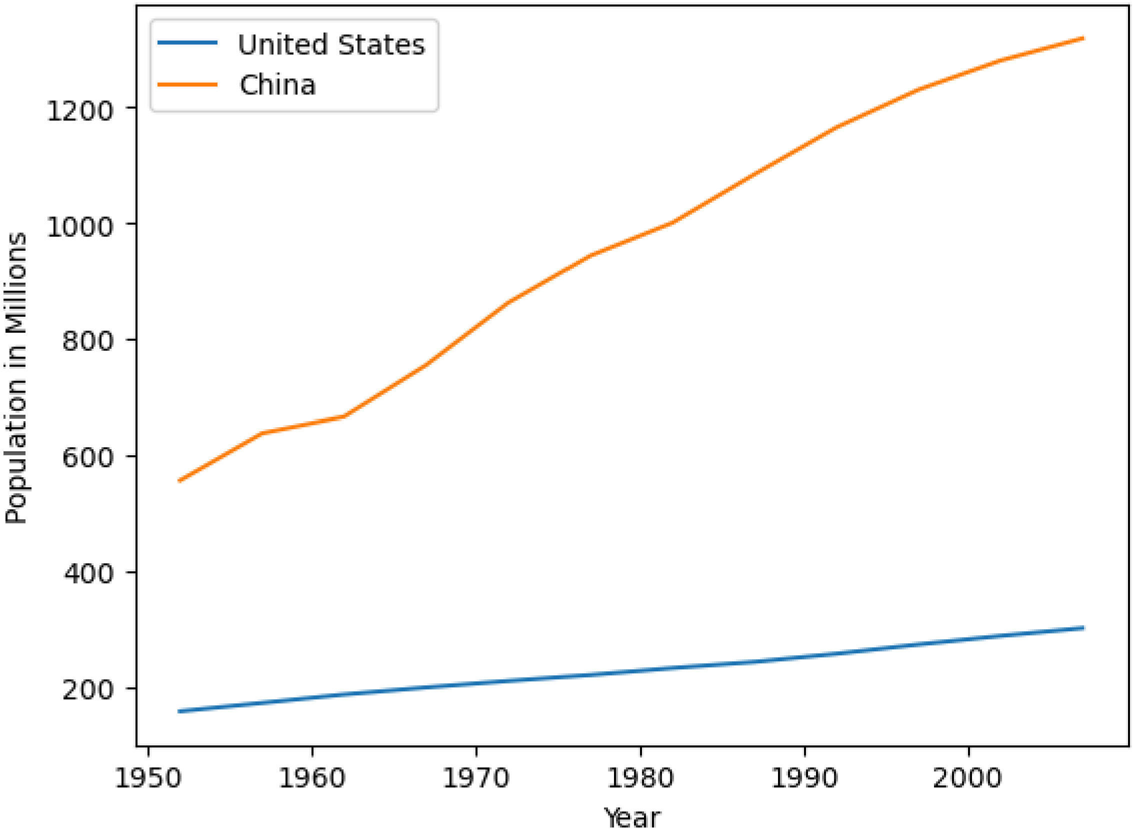


In [22]:
```
plt.plot(us.year, us.population / 10**6) # divide the population by 1 million
plt.plot(china.year, china.population / 10**6) # divide the population by 1 millior
plt.show ()
```



In [23]:

```
plt.plot(us.year, us.population / 10**6) # divide the population by 1 million
plt.plot(china.year, china.population / 10**6) # divide the population by 1 millior
plt.legend(['United States', 'China'])
plt.xlabel('Year')
plt.ylabel('Population in Millions')
plt.show ()
```



In [24]:

```
us.population
```

Out[24]:

```
1608    157553000
1609    171984000
1610    186538000
1611    198712000
1612    209896000
1613    220239000
1614    232187835
1615    242803533
1616    256894189
1617    272911760
1618    287675526
1619    301139947
Name: population, dtype: int64
```

In [25]:

```
us.population.iloc[0]
```

Out[25]:

```
157553000
```

In [26]:

```
us.population / us.population.iloc[0] *100
```

Out[26]:

```
1608    100.000000
1609    109.159457
1610    118.396984
1611    126.123908
1612    133.222471
1613    139.787246
1614    147.371256
1615    154.109114
1616    163.052553
1617    173.219018
1618    182.589685
1619    191.135648
Name: population, dtype: float64
```

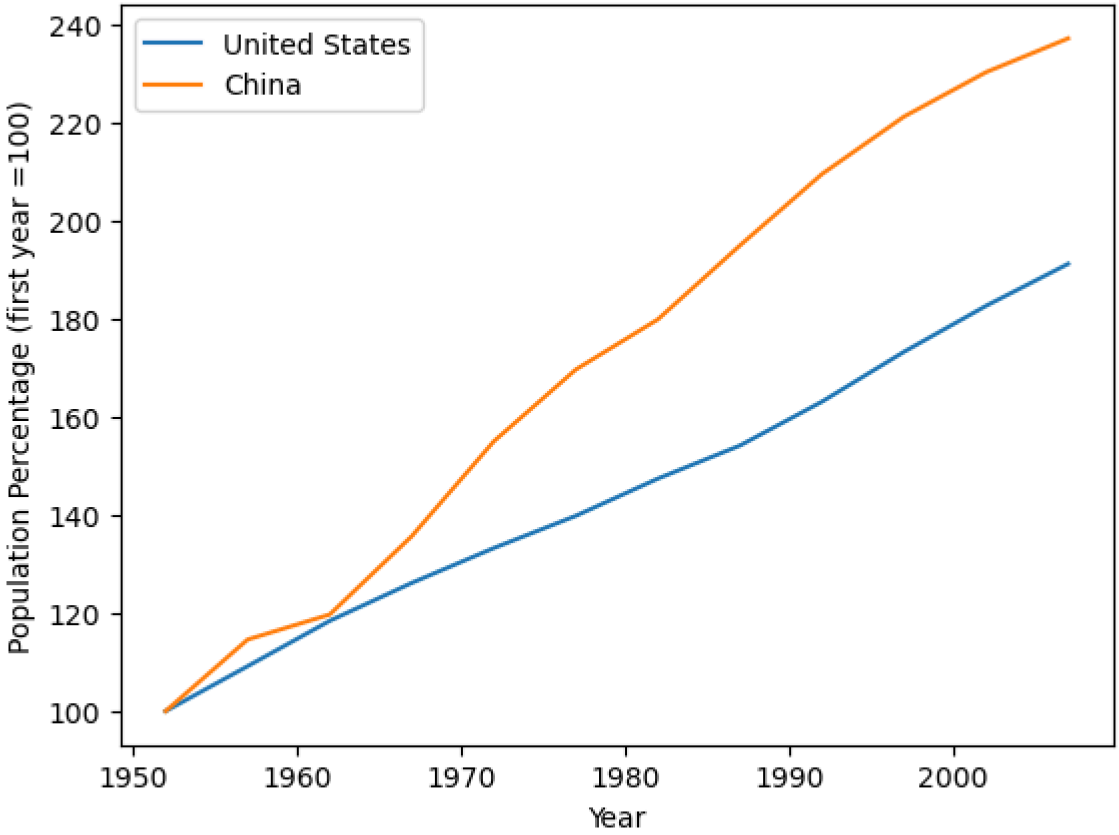
In [ ]:

```

```

In [27]:

```
plt.plot(us.year,us.population / us.population.iloc[0] *100) # divide the populatio
plt.plot(china.year, china.population / china.population.iloc[0] *100) # divide the
plt.legend(['United States', 'China'])
plt.xlabel('Year')
plt.ylabel('Population Percentage (first year =100)')
plt.show ()
```



find the percentage of population increase in United States and China for the year 2007

(Source File: countries.csv)

In [32]:
```
us.year==2007
```

Out[32]:
```
1608    False
1609    False
1610    False
1611    False
1612    False
1613    False
1614    False
1615    False
1616    False
1617    False
1618    False
1619     True
Name: year, dtype: bool
```

In [45]:
```
population_usa_2007=us.population[us.year==2007]
print(population_usa_2007)
```

```
1619    301139947
Name: population, dtype: int64
```

In [48]:
```
print(population_usa_2007 / us.population.iloc[0] *100)
```

```
1619    191.135648
Name: population, dtype: float64
```

In [ ]: