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

```
import pandas as pd # We'll be using Pandas library to work with the dataset
from matplotlib import pyplot as plt

data=pd.read_csv('countries.csv') #read the file as data

data # displays the data set
```

```
          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 x 3 columns]
```

```
type(data) #type of data
```

```
pandas.core.frame.DataFrame
```

```
data.tail() # the tail command to see the last 5 items in the csv file
```

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

```
data.info() # we'll be able to see all of the available columns in the dataset along with their corresponding data types.
```

```
<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
```

```
data.describe() # the index results include the count, mean, std,
minimum 25%, 50%, 75%, and maximum from the dataset.
```

```
              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
```

```
#compare the population of US and China

#isolate the data of US and China

data.country == 'United States'#showcase when and where USA as True
```

```
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
```

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

us
```

```
            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
```

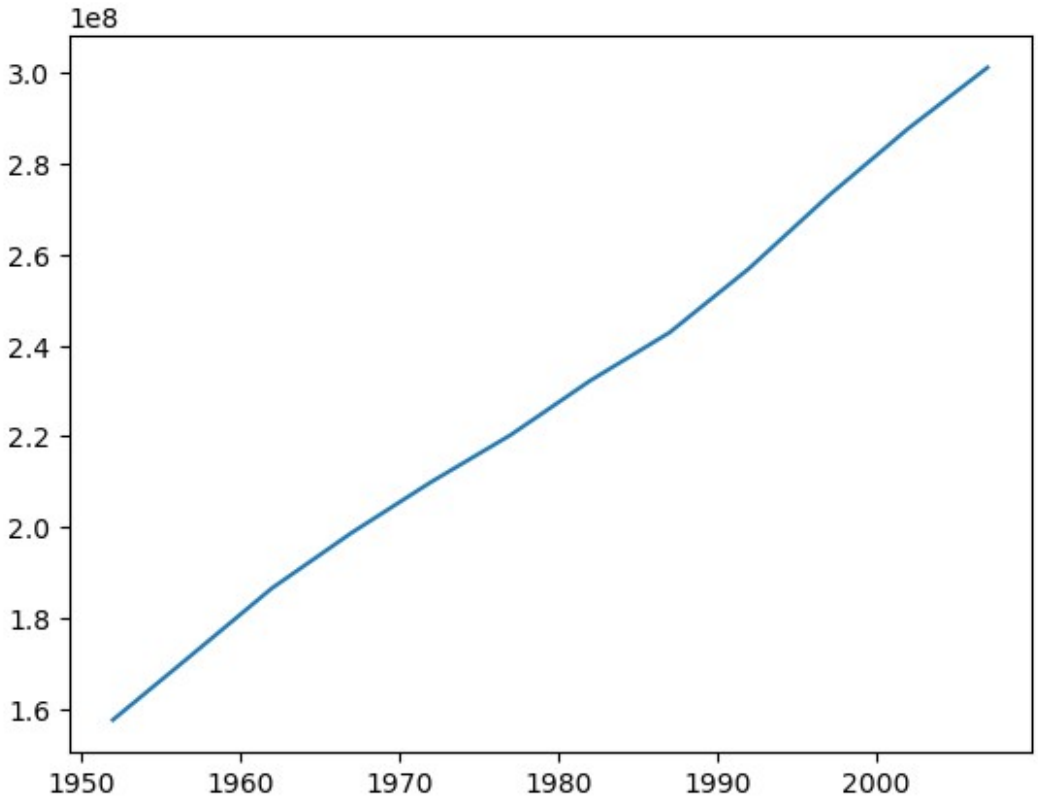
```
1615  United States  1987   242803533
1616  United States  1992   256894189
1617  United States  1997   272911760
1618  United States  2002   287675526
1619  United States  2007   301139947
```

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

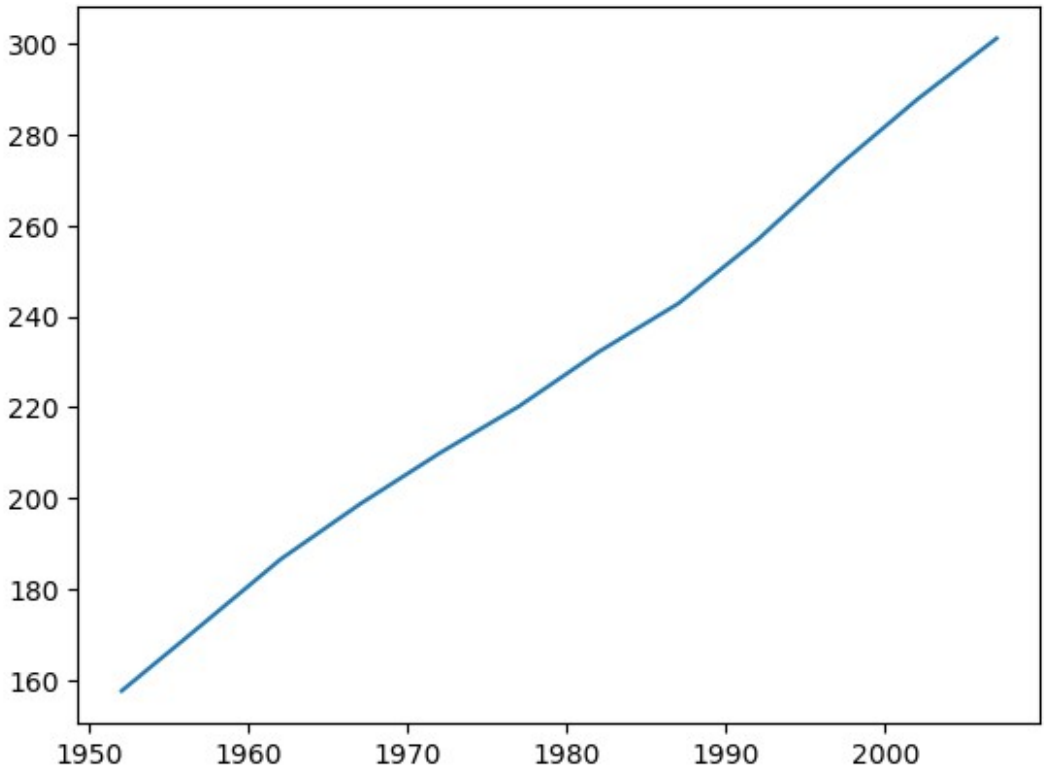
```
china
```

```
    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
```

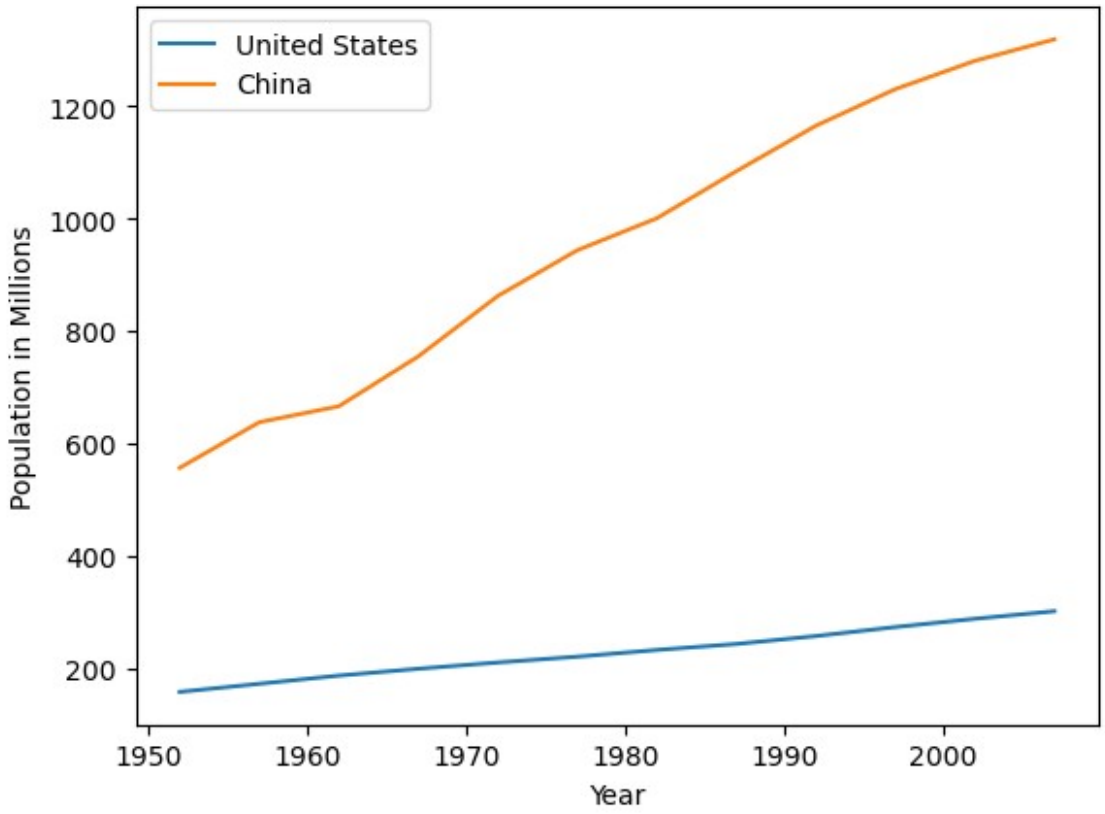
```
plt.plot(us.year, us.population)
plt.show ()
```




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




```
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 million
plt.legend(['United States', 'China'])
plt.xlabel('Year')
plt.ylabel('Population in Millions')
plt.show ()
```




```
us.population
```

```
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
```

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

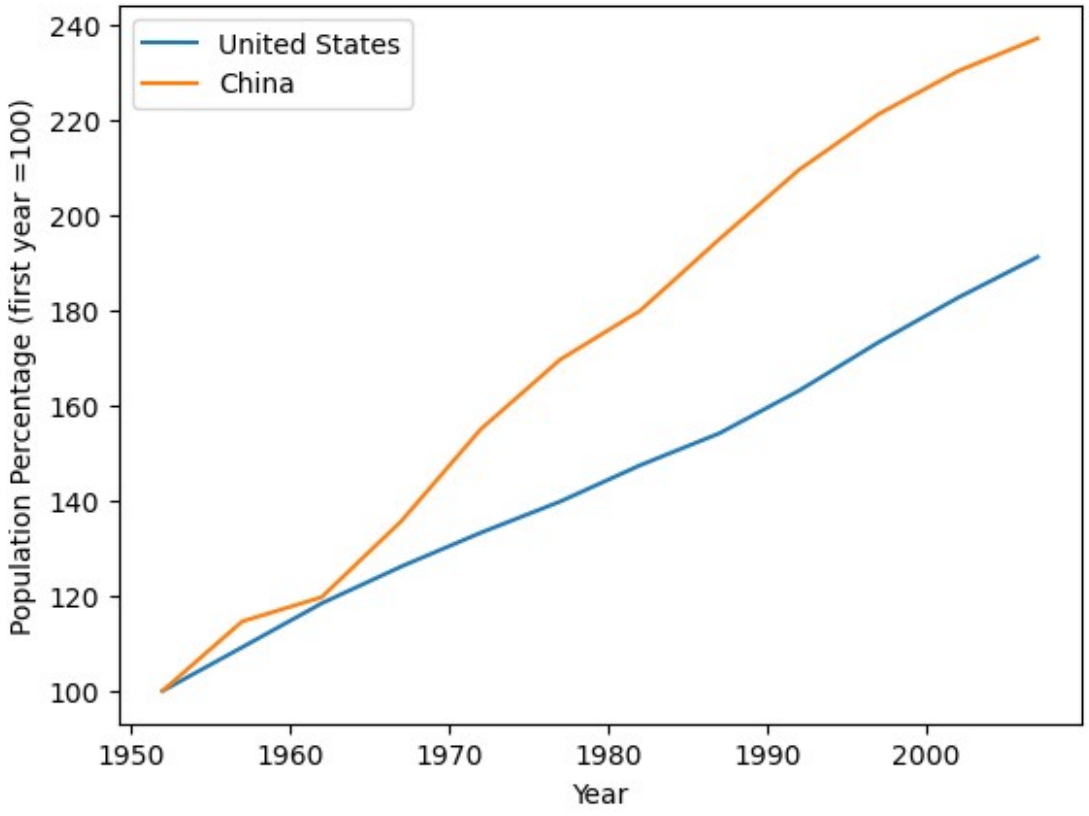
```
157553000
```

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

```
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
```

```
plt.plot(us.year,us.population / us.population.iloc[0] *100) # divide
the population by 1 million
plt.plot(china.year, china.population / china.population.iloc[0] *100)
# divide the population by 1 million
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)

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

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

```
print(population_usa_2007 / us.population.iloc[0] *100)  # % growth for USA
```

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

```
population_china_2007=china.population[china.year==2007]
print(population_china_2007)
```

```
299    1318683096
Name: population, dtype: int64
```

```
print(population_china_2007 / china.population.iloc[0] *100)  # % growth for china
```

```
299    237.060859
Name: population, dtype: float64
```

```
# Assignment : find the percentage of population increase in India for the year 2002
```