

Population

811_Suraj Didwagh
816_Ojas Gharde
810_Harshvardhan Dhoot
817_Pratik Gole

Code :

```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
from sklearn.linear_model import LinearRegression
import seaborn as sns
```

```
fp=open("population 2.csv",'r')
data=fp.read()
lines=data.splitlines()
```

```
china=[]
india=[]
usa=[]
yearc=[]
yeari=[]
yearu=[]
country=['China','India','USA']
```

```
for line in lines:
    word=line.split(',')
    if(word[1]=="CHN"):
        china.append(str(word[3]))
        yearc.append(str(word[2]))
    elif(word[1]=="IND"):
        india.append(str(word[3]))
        yeari.append(str(word[2]))
    else:
        usa.append(str(word[3]))
        yearu.append(str(word[2]))

plt.plot(china,yearc)
plt.grid()
plt.ylabel("Years")
plt.xlabel("Population")
plt.show()

plt.plot(india,yeari)
plt.grid()
plt.ylabel("Years")
plt.xlabel("Population")
plt.show()
```

```
plt.plot(usa,yearu)
plt.grid()
plt.ylabel("Years")
plt.xlabel("Population")
plt.show()
```

```
fp=pd.read_csv('population 2.csv')
```

```
pop=fp.groupby('Country')['Population'].sum().idxmax()
print("The country with maximum pupulation is: ",pop)
```

```
mpop=fp.groupby('Country')['Population'].max()
print("The maximum population of each country(in billion) is:
\n",mpop)
```

```
popn=fp.groupby('Country')['Population'].sum().idxmin()
print("The country with minimum pupulation is: ",popn)
```

```
npop=fp.groupby('Country')['Population'].min()
print("The minimum population of each country(in billion) is: \n",npop)
```

```
meanpop=fp.groupby('Country')['Population'].mean()
print("The mean population of each country is: \n",meanpop)
```

```
plt.bar(country,meanpop)
plt.show()
```

```
plt.bar(country,mpop)
plt.show()
```

```
plt.bar(country,npop)
plt.show()
```

```
data2 = {
    'Country': ['China']*11 + ['India']*11 + ['USA']*11,
    'Year': [2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021]*3,
    'Population(cr)': [135, 136, 137, 138, 139, 140, 141, 141, 142, 142, 142,
                       125, 127, 129, 130, 132, 133, 135, 136, 138, 139, 140,
                       31.3, 31.6, 31.9, 32.2, 32.4, 32.7, 32.9, 33.2, 33.4, 33.5, 33.6]
}
```

```
df = pd.DataFrame(data2)
print(df.head())
print(df.describe())
avg_population = df.groupby('Country')['Population(cr)'].mean()
print(avg_population)
```

```
X = df['Year'].values.reshape(-1, 1)
y = df['Population(cr)'].values
```

```
model = LinearRegression()
model.fit(X, y)
```

```
population_2022 = model.predict([[2022]])
print("Predicted population in 2022:", population_2022)
correlation = df['Year'].corr(df['Population(cr)'])
print("Correlation between Year and Population:", correlation)
max_population_country = df[df['Year'] == 2021].sort_values('Population(cr)',
ascending=False).iloc[0]['Country']
print("Country with the highest population in 2022:", max_population_country)
```

```
sns.lineplot(data=df, x='Year', y='Population(cr)', hue='Country')
```

```
plt.title('Population Over Time')  
plt.xlabel('Year')  
plt.ylabel('Population (in crores)')  
plt.show()
```

```
avg_population = df.groupby('Country')['Population(cr)'].mean().reset_index()
```

```
sns.barplot(data=avg_population, x='Country', y='Population(cr)')  
plt.title('Average Population by Country')  
plt.xlabel('Country')
```

```
sns.scatterplot(data=df, x='Year', y='Population(cr)', hue='Country')  
plt.title('Year vs. Population')  
plt.xlabel('Year')  
plt.ylabel('Population (in crores)')  
plt.show()
```

```
sns.boxplot(data=df, x='Country', y='Population(cr)')  
plt.title('Population by Country')  
plt.xlabel('Country')  
plt.ylabel('Population (in crores)')  
plt.show()
```

```
corr_matrix = df.corr()
sns.heatmap(corr_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()
```

```
X = df[['Year']]
y = df['Population(cr)']
model = LinearRegression()
model.fit(X,y)
future_years = pd.DataFrame({'Year': [2022, 2023, 2024]})
population_predictions = model.predict(future_years)
future_years = pd.DataFrame({'Year': [2022, 2023, 2024]})
population_predictions = model.predict(future_years)
plt.scatter(df['Year'], df['Population(cr)'], label='Actual Data')
plt.plot(future_years, population_predictions, color='red', label='Predicted
Data')
plt.title('Linear Regression - Population Prediction')
plt.xlabel('Year')
plt.ylabel('Population (in crores)')
plt.legend()
plt.show()
```



```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
from sklearn.linear_model import LinearRegression
import seaborn as sns
```

```
fp=open("population 2.csv",'r')
data=fp.read()
lines=data.splitlines()
print
```

```
china=[]
india=[]
usa=[]
yearc=[]
yeari=[]
yearu=[]
country=['China','India','USA']
```

```
for line in lines:
    word=line.split(',')
    if(word[1]=="CHN"):
        china.append(str(word[3]))
        yearc.append(str(word[2]))
    elif(word[1]=="IND"):
        india.append(str(word[3]))
        yeari.append(str(word[2]))
    else:
        usa.append(str(word[3]))
        yearu.append(str(word[2]))
plt.plot(china,yearc)
plt.grid()
plt.ylabel("Years")
plt.xlabel("Population")
plt.show()
plt.plot(india,yeari)
plt.grid()
plt.ylabel("Years")
plt.xlabel("Population")
plt.show()
```

```
plt.plot(usa,yearu)
plt.grid()
plt.ylabel("Years")
plt.xlabel("Population")
plt.show()
```

```
fp=pd.read_csv('population 2.csv')
```

```
pop=fp.groupby('Country')['Population'].sum().idxmax()
print("The country with maximum pupulation is: ",pop)
```

```
mpop=fp.groupby('Country')['Population'].max()
print("The maximum population of each country(in billion) is: \n",mpop)
```

```
popn=fp.groupby('Country')['Population'].sum().idxmin()
print("The country with minimum pupulation is: ",popn)
```

```
npop=fp.groupby('Country')['Population'].min()  
print("The minimum population of each country(in  
billion) is: \n",npop)
```

```
meanpop=fp.groupby('Country')['Population'].mean()  
print("The mean population of each country is:  
\n",meanpop)
```

Output :

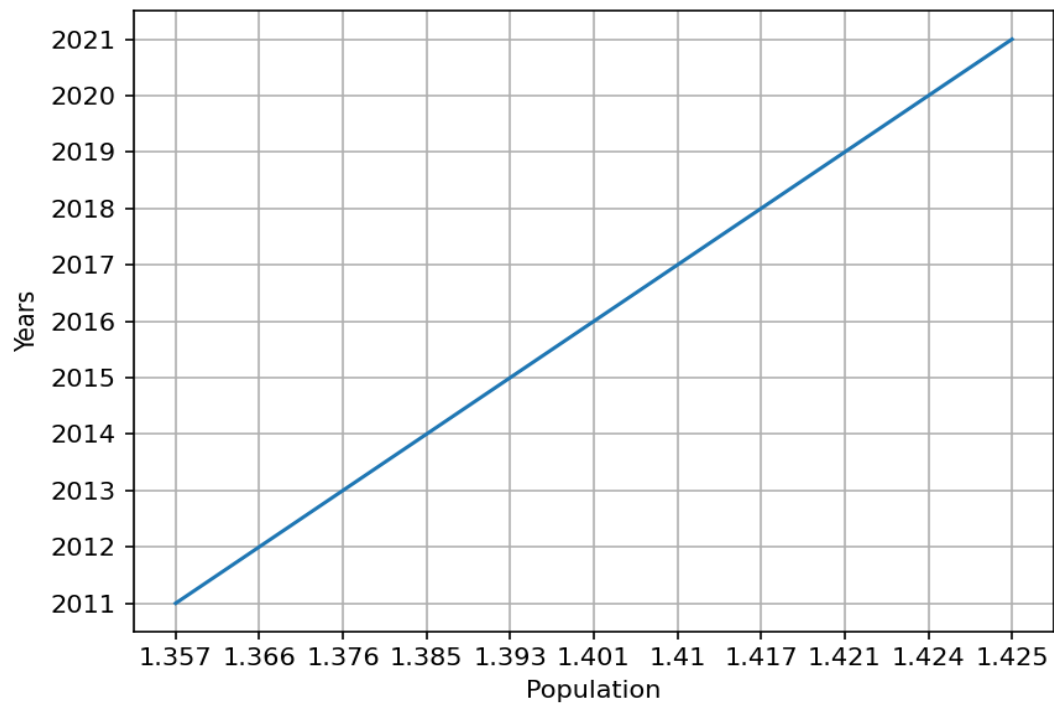
```
===== RESTART: C:/Users/pgole/Desktop/EDS PROJECT/project.py =====
The country with maximum pupulation is: China
The maximum population of each country(in billion) is:
Country
China          1.425
India           1.407
United States   0.336
Name: Population, dtype: float64
The country with minimum pupulation is: United States
The minimum population of each country(in billion) is:
Country
China          1.357
India           1.257
United States   0.313
Name: Population, dtype: float64
The mean population of each country is:
Country
China          1.397727
India           1.336182
United States   0.326091
Name: Population, dtype: float64
  Country  Year  Population(cr)
0  China  2011           135.0
1  China  2012           136.0
2  China  2013           137.0
3  China  2014           138.0
4  China  2015           139.0
      Year  Population(cr)
count    33.000000      33.000000
mean    2016.000000    101.687879
std       3.211308      49.772029
min    2011.000000      31.300000
25%    2013.000000      33.400000
50%    2016.000000     133.000000
75%    2019.000000     139.000000
max    2021.000000     142.000000
```

```

4      China      2019      139.0
      Year  Population(cr)
count      33.000000      33.000000
mean    2016.000000    101.687879
std        3.211308    49.772029
min    2011.000000    31.300000
25%    2013.000000    33.400000
50%    2016.000000    133.000000
75%    2019.000000    139.000000
max    2021.000000    142.000000
Country
China    139.363636
India    133.090909
USA       32.609091
Name: Population(cr), dtype: float64
Predicted population in 2022: [106.65333333]
Correlation between Year and Population: 0.05339546801656071
Country with the highest population in 2022: China

```

Figure 1



x=1.421 y=2012

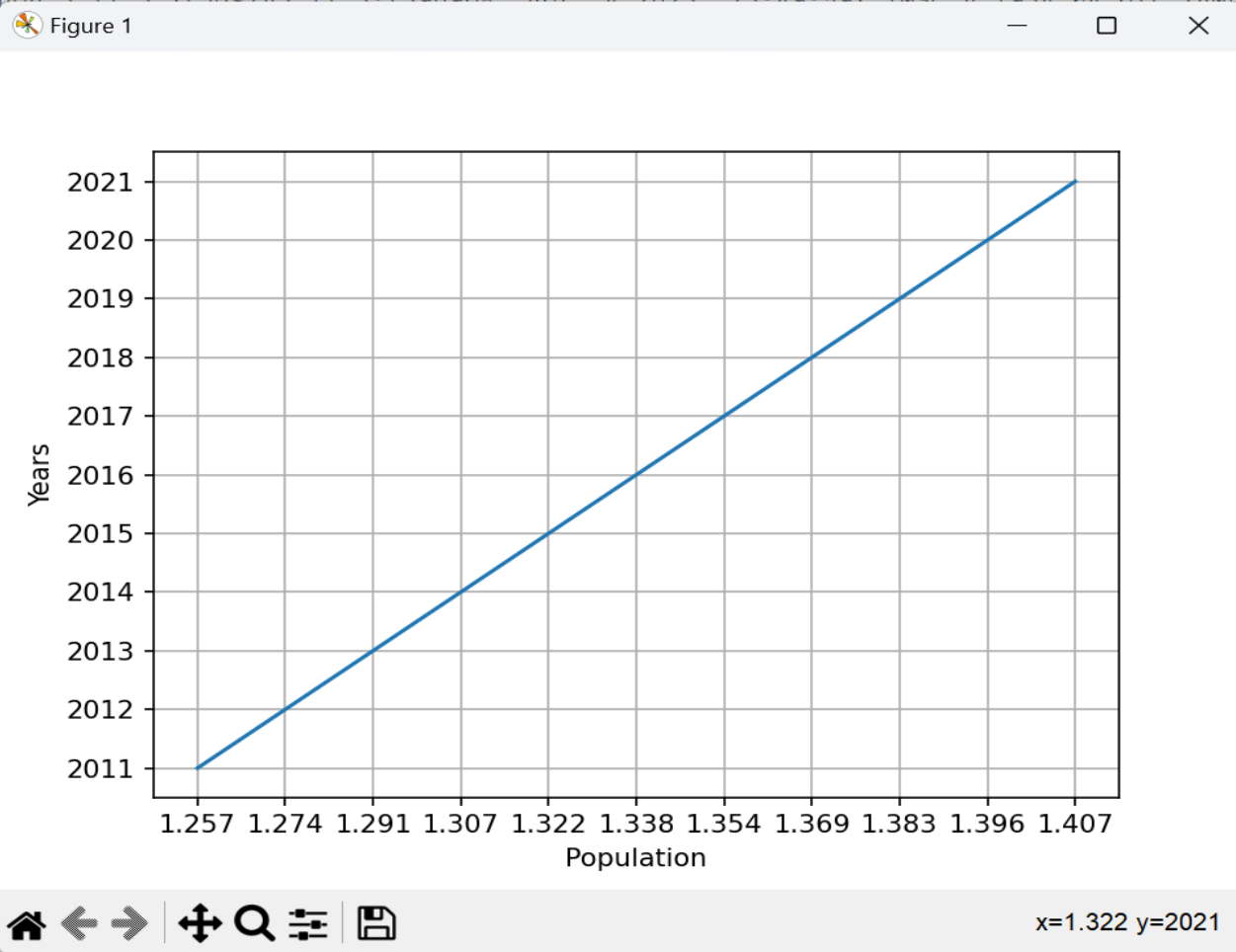
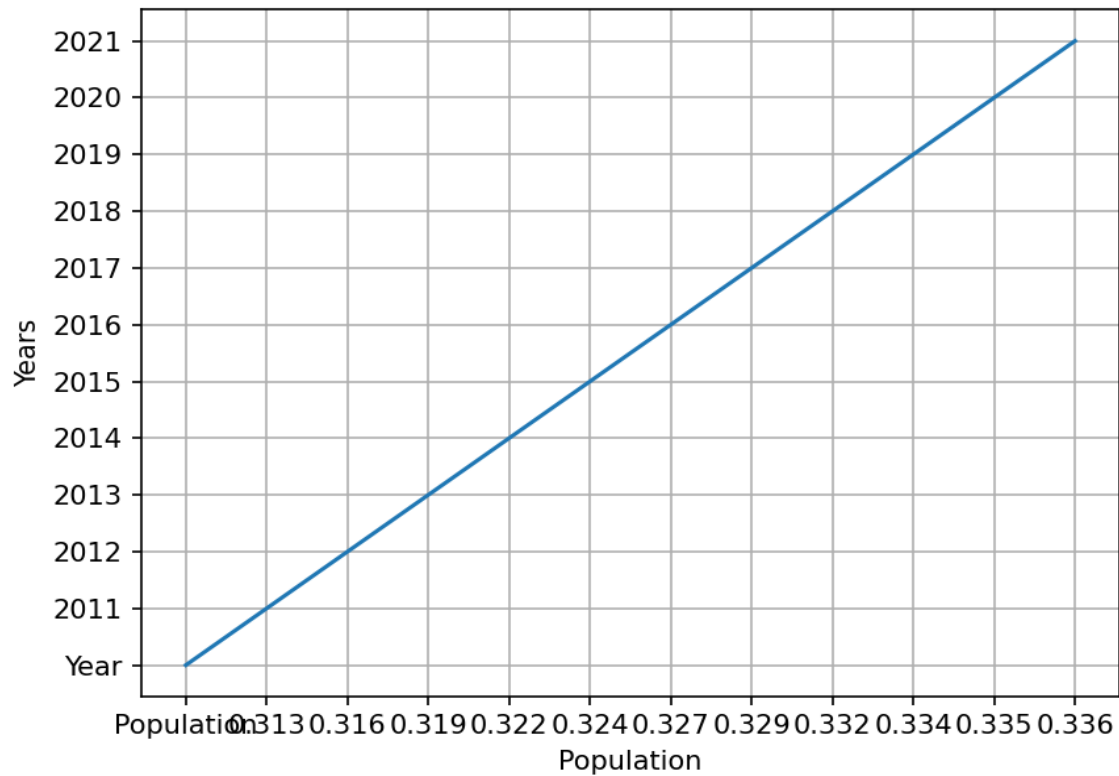


Figure 1



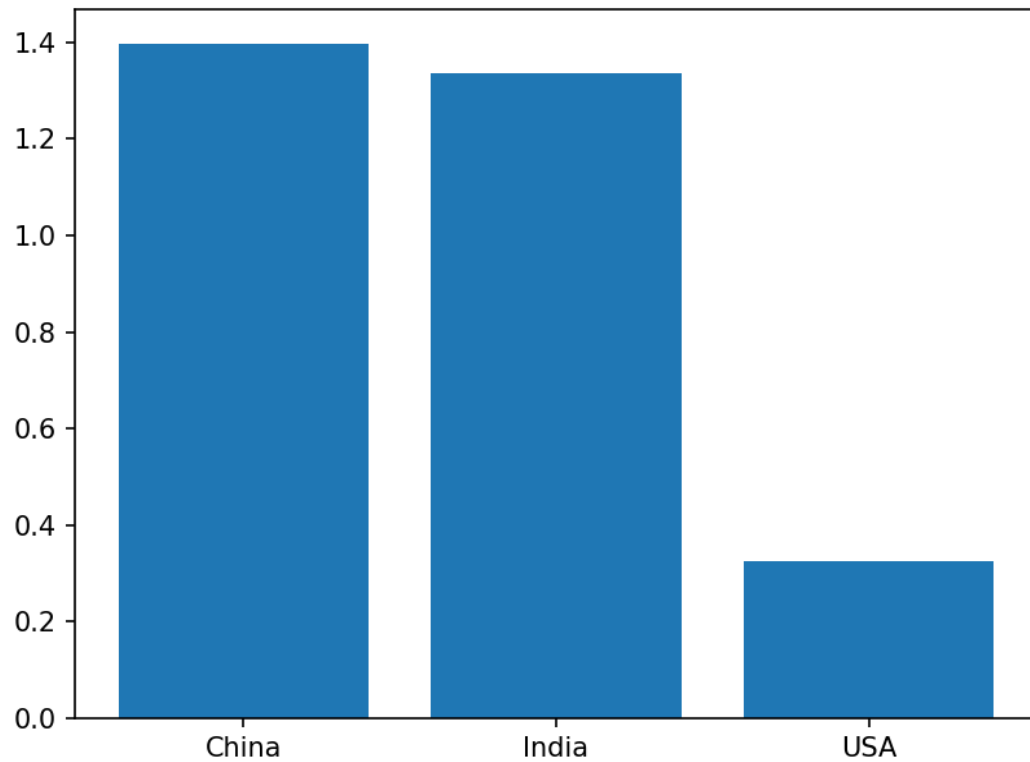


Figure 1

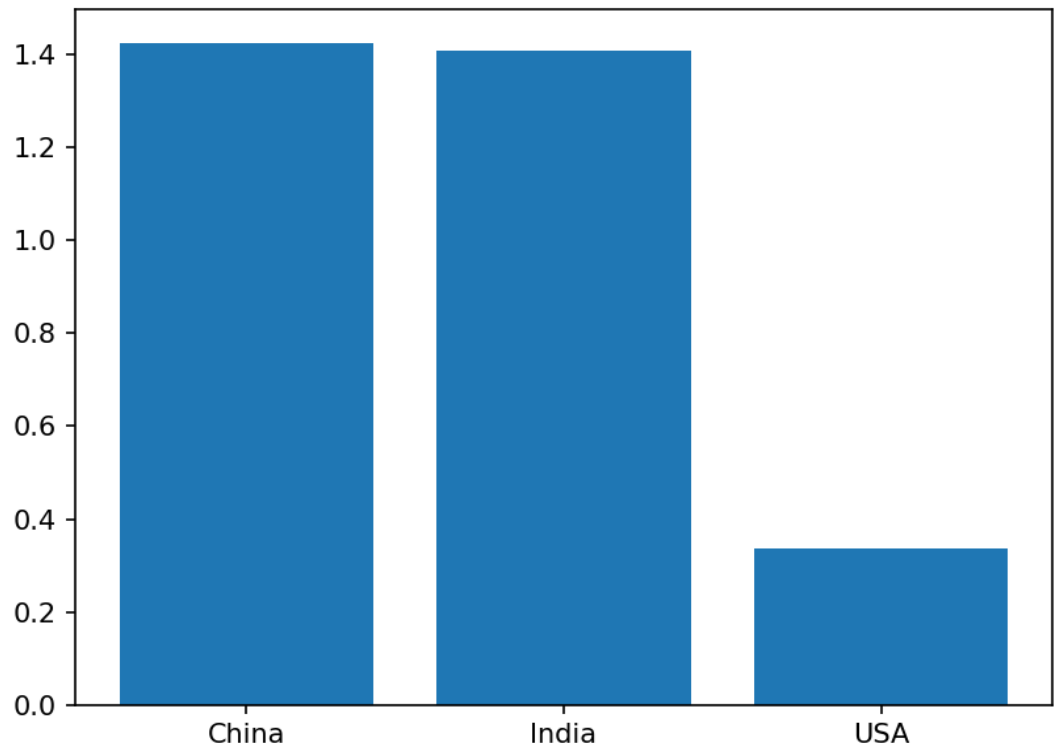
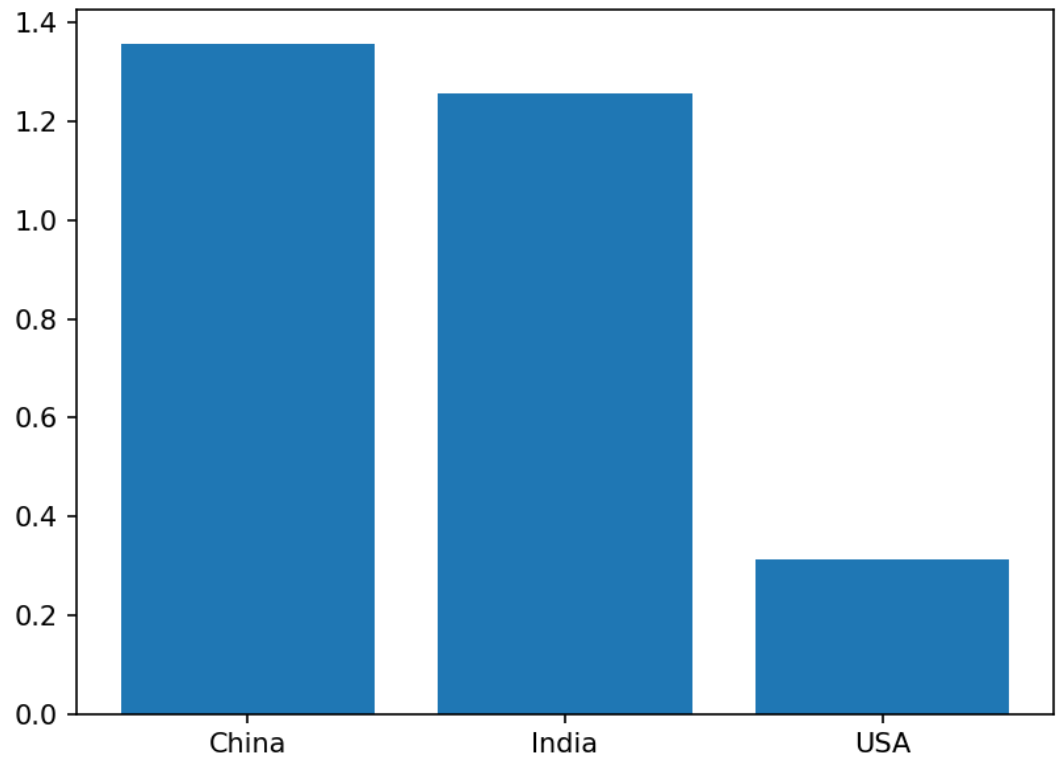


Figure 1



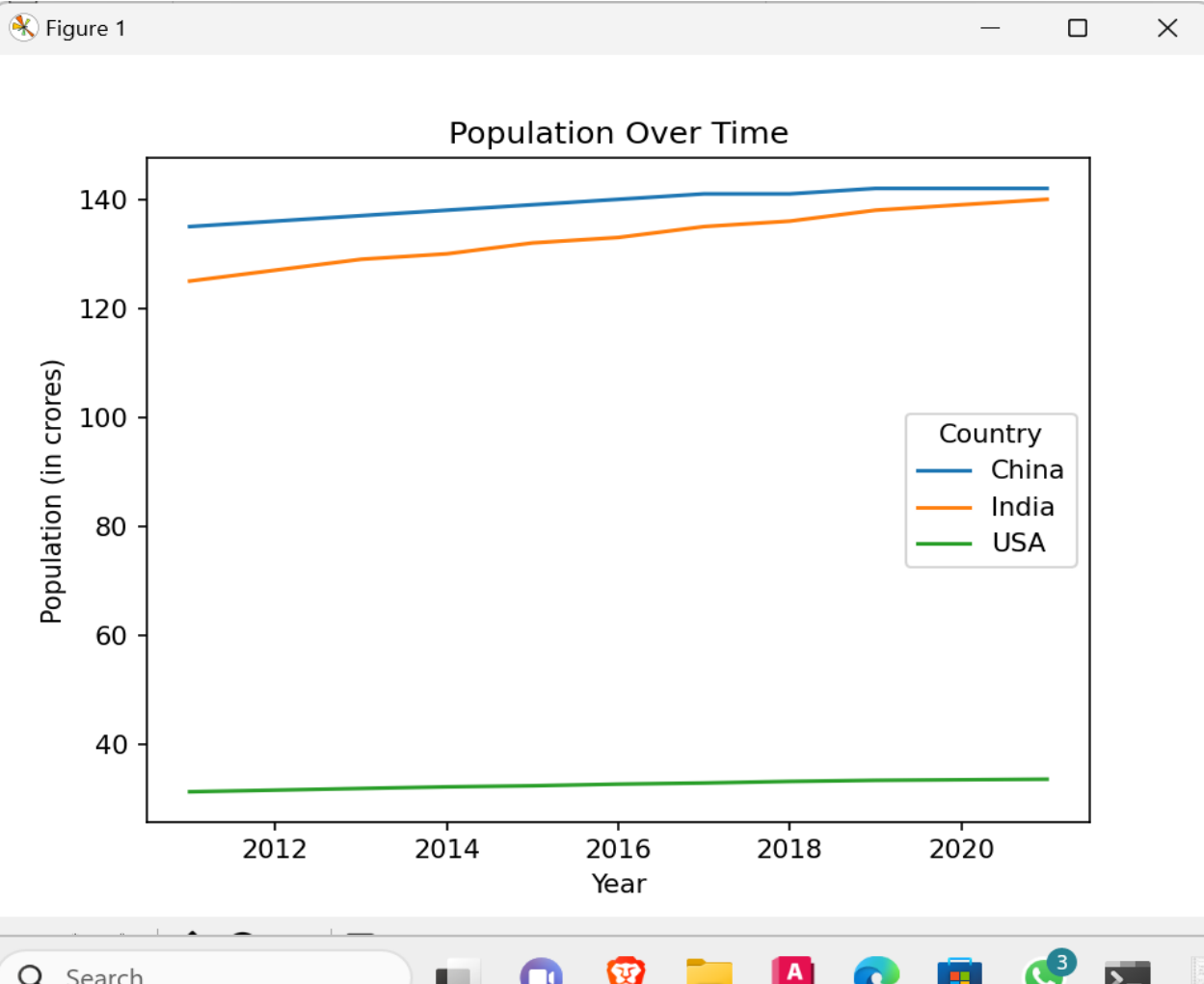


Figure 1

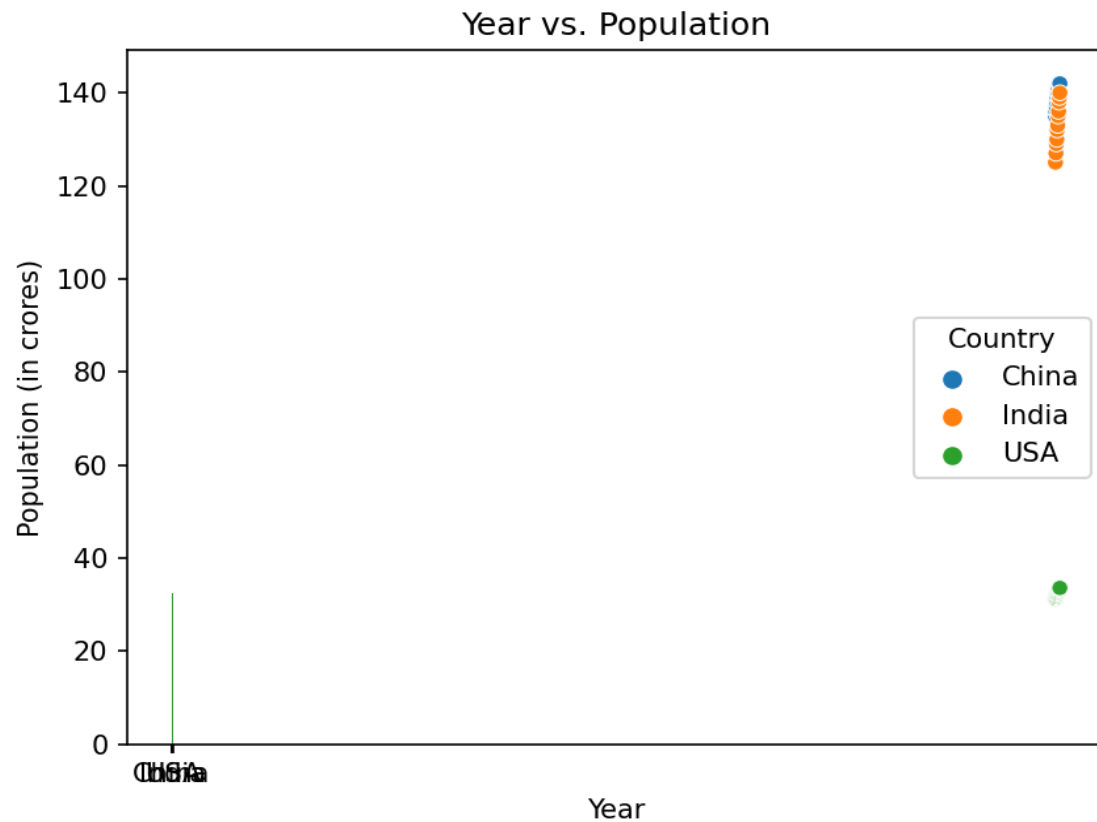


Figure 1

