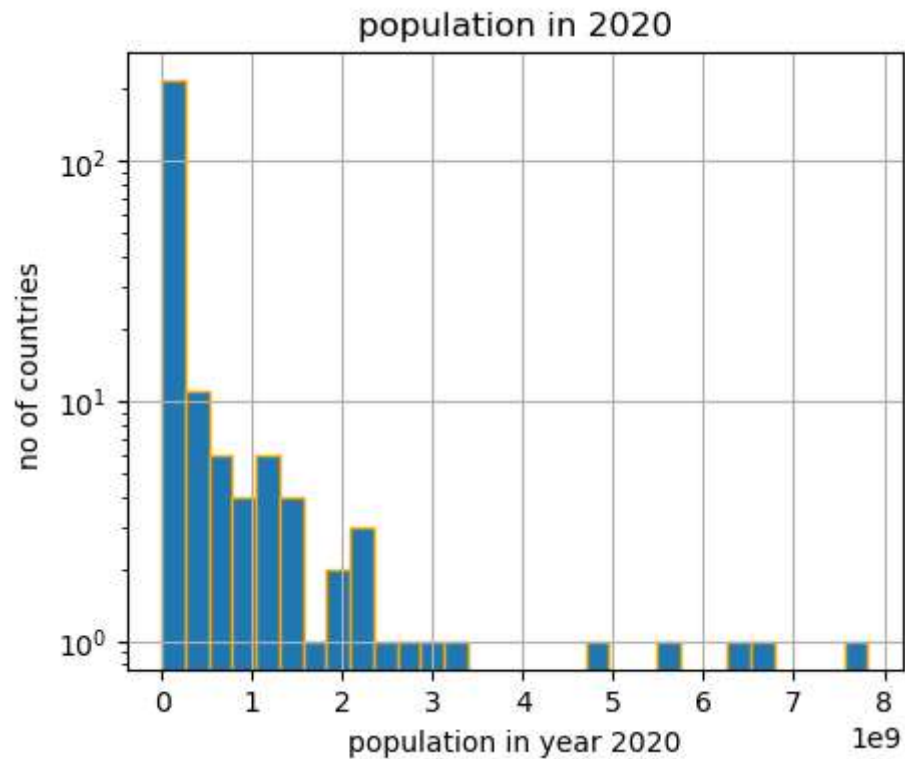


```
In [1]: # import required libraries for the process  
import pandas as pd  
import matplotlib.pyplot as plt
```

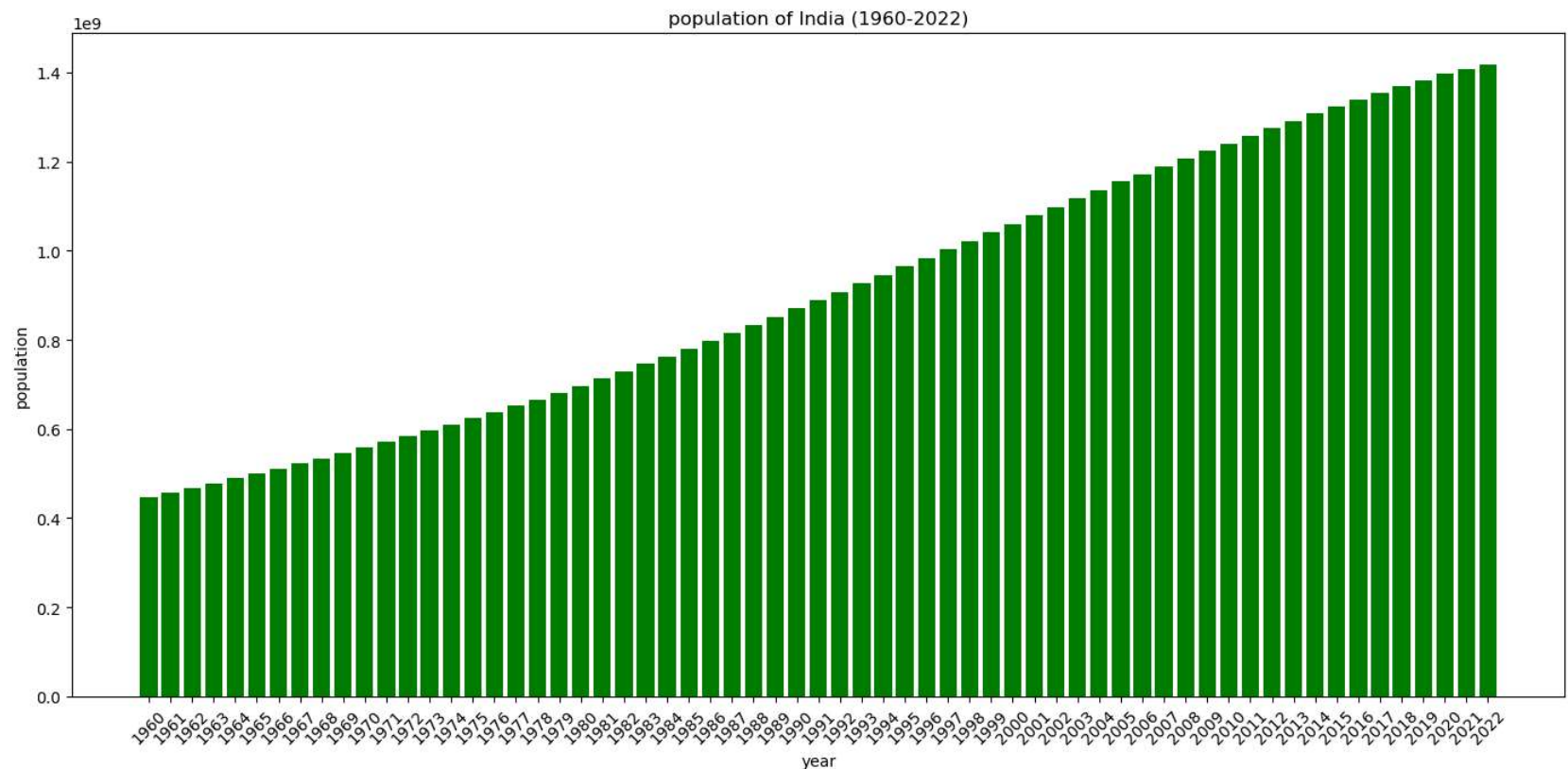
```
In [2]: # import the dataset from the form of csv  
df = pd.read_csv("dataset_task1.csv",skiprows=4)
```

```
In [3]: population_2020 = df['2020'].dropna()
```

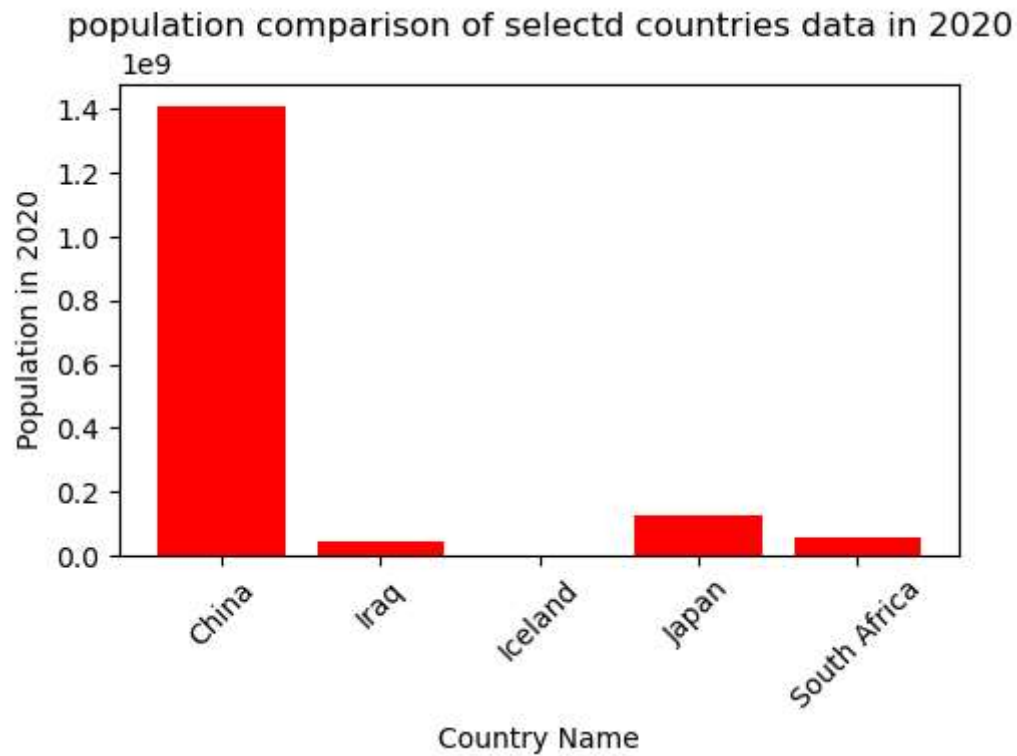
```
In [13]: # plot the population in no of counties
plt.figure(figsize=(5,4))
plt.hist(population_2020,bins=30,edgecolor='orange')
plt.title('population in 2020')
plt.xlabel('population in year 2020')
plt.ylabel('no of countries ')
plt.yscale('log')
plt.grid(True)
plt.show()
```



```
In [24]: # plot afghanistan yearwise
India_data = df[df['Country Name'] == 'India'].iloc[0,4:-2]
India_data = pd.to_numeric(India_data,errors='coerce').dropna()
plt.figure(figsize=(14,7))
plt.bar(India_data.index,India_data.values,color='green')
plt.xlabel('year')
plt.ylabel('population')
plt.title('population of India (1960-2022)')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



```
In [25]: countries = ['Iraq', 'Iceland', 'Japan', 'China', 'South Africa']
population_2020 = df[df['Country Name'].isin(countries)][['Country Name', '2020']]
plt.figure(figsize=(5,4))
plt.bar(population_2020['Country Name'],population_2020['2020'],color='red')
plt.xlabel('Country Name')
plt.ylabel('Population in 2020')
plt.title('population comparison of selectd countries data in 2020')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



```
In [26]: population_data = df[df['Country Name'].isin(countries)].set_index('Country Name')

years = population_data.columns[4:-2]

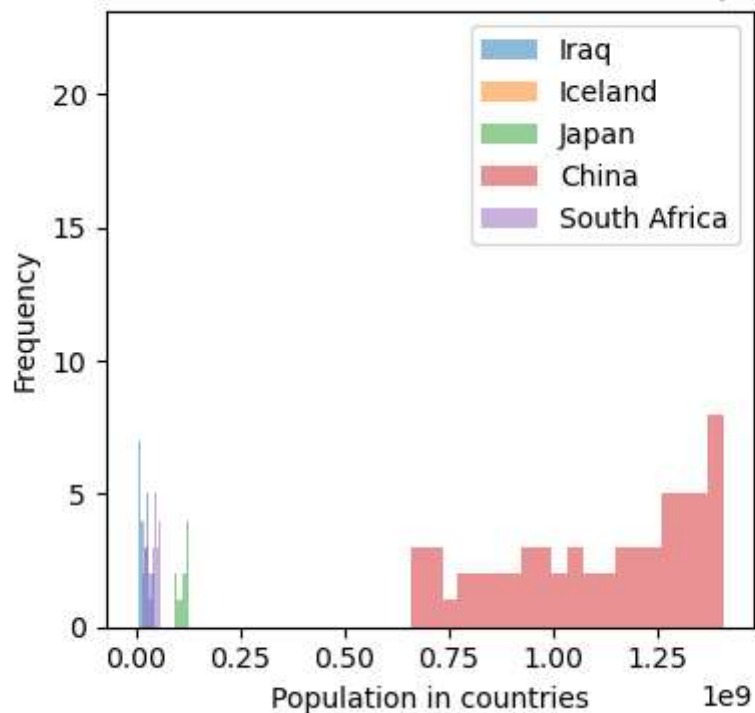
plt.figure(figsize=(4, 4))

for country in countries:
    country_data = pd.to_numeric(population_data.loc[country, years], errors='coerce').dropna()
    plt.hist(country_data.values, bins=20, alpha=0.5, label=country)

plt.xlabel('Population in countries')
plt.ylabel('Frequency')
plt.title('Population Distribution for Selected Countries (1960-2022)')
plt.legend()
plt.tight_layout()

plt.show()
```

Population Distribution for Selected Countries (1960-2022)



```
In [27]: region = 'Africa Eastern and Southern'
region_data = df[df['Country Name'] == region].iloc[0, 4:-2]

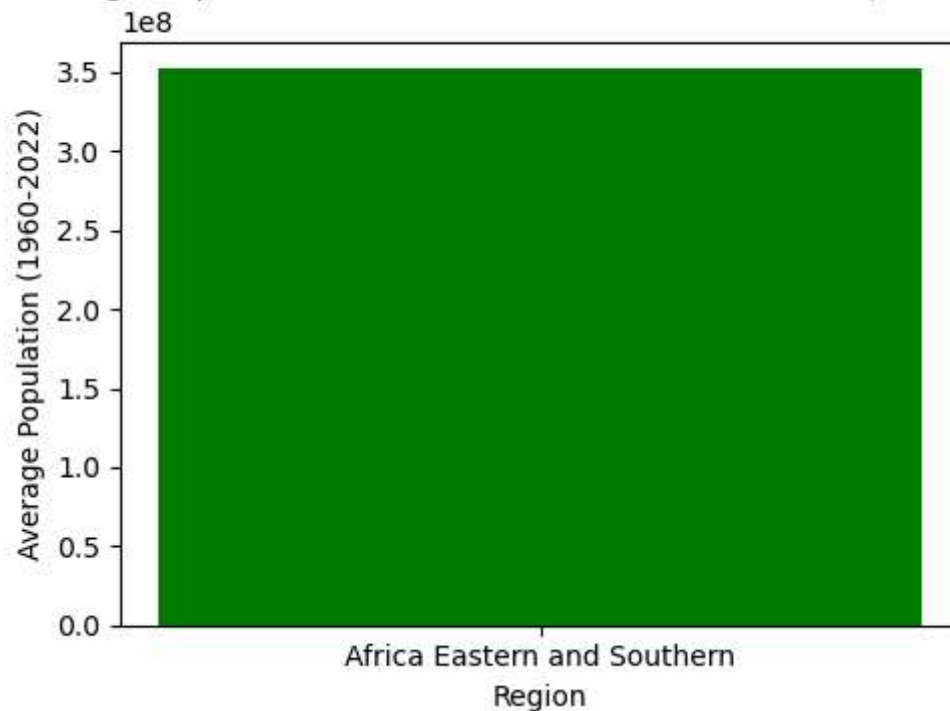
region_data = pd.to_numeric(region_data, errors='coerce').dropna()

average_population = region_data.mean()

plt.figure(figsize=(5, 4))
plt.bar([region], [average_population], color='green')
plt.xlabel('Region')
plt.ylabel('Average Population (1960-2022)')
plt.title('Average Population of Africa Eastern and Southern (1960-2022)')
plt.tight_layout()

plt.show()
```

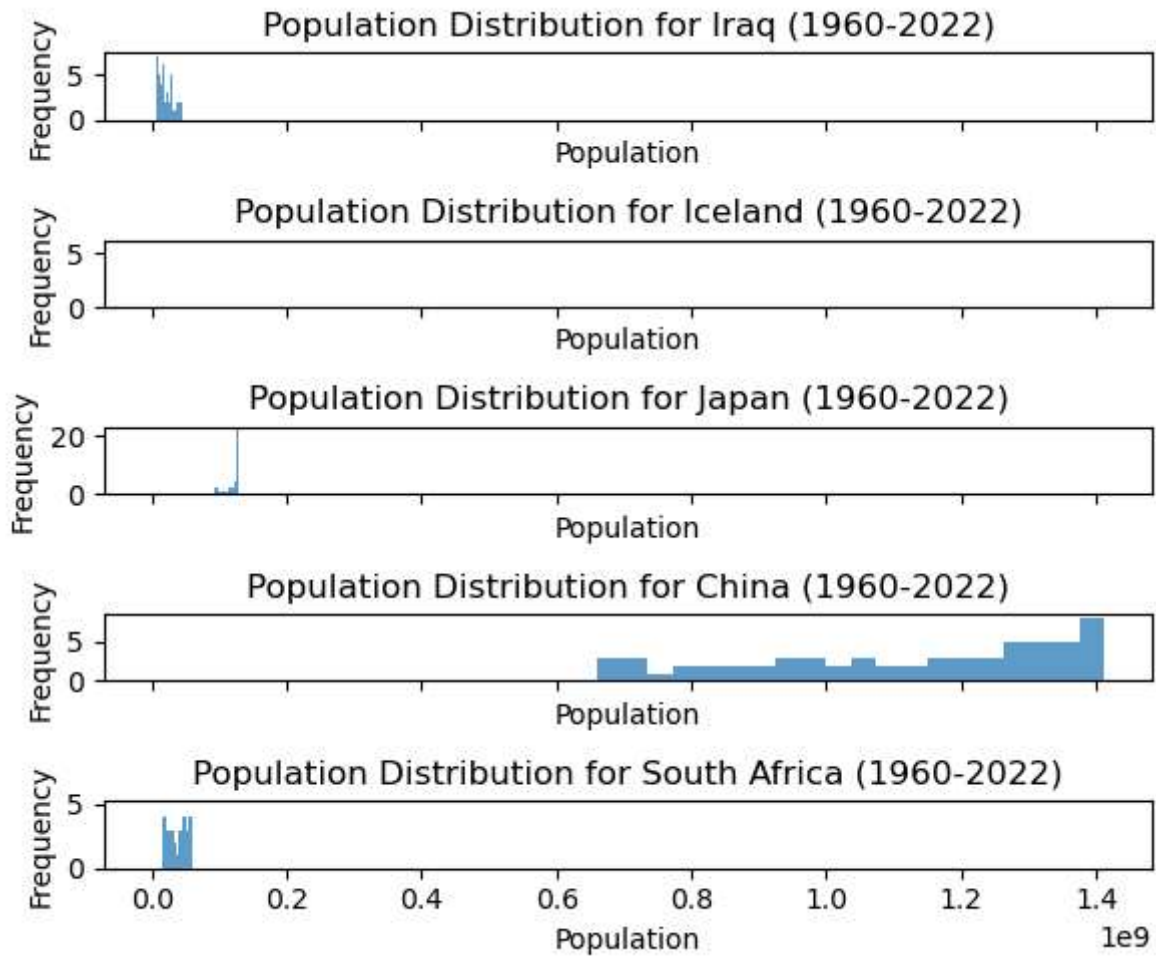
Average Population of Africa Eastern and Southern (1960-2022)



```
In [28]: population_data = df[df['Country Name'].isin(countries)].set_index('Country Name')
years = population_data.columns[4:-2]

fig, axes = plt.subplots(len(countries), 1, figsize=(6, 5), sharex=True)
for i, country in enumerate(countries):
    country_data = pd.to_numeric(population_data.loc[country, years], errors='coerce').dropna()
    axes[i].hist(country_data.values, bins=20, alpha=0.7)
    axes[i].set_title(f'Population Distribution for {country} (1960-2022)')
    axes[i].set_xlabel('Population')
    axes[i].set_ylabel('Frequency')

plt.tight_layout()
plt.show()
```



In []: