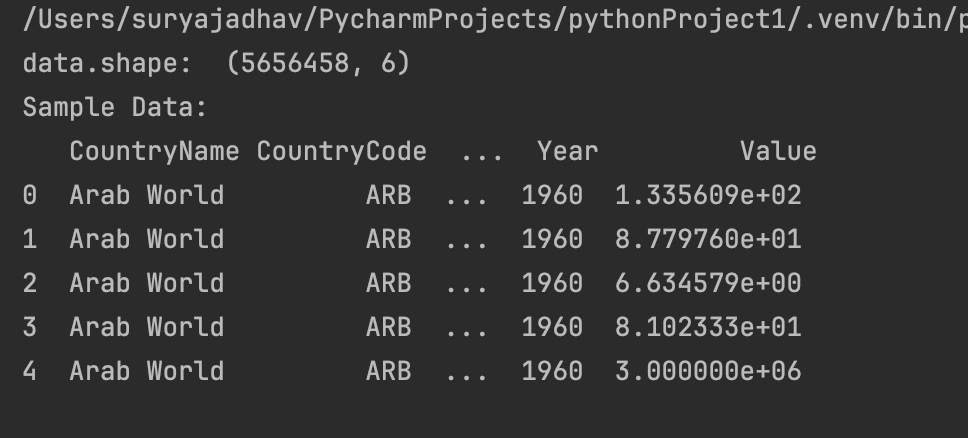
PROJECT 2: VISUALIZATION OF WORLD GDP AND CARBON - DIOXIDE EMISSION

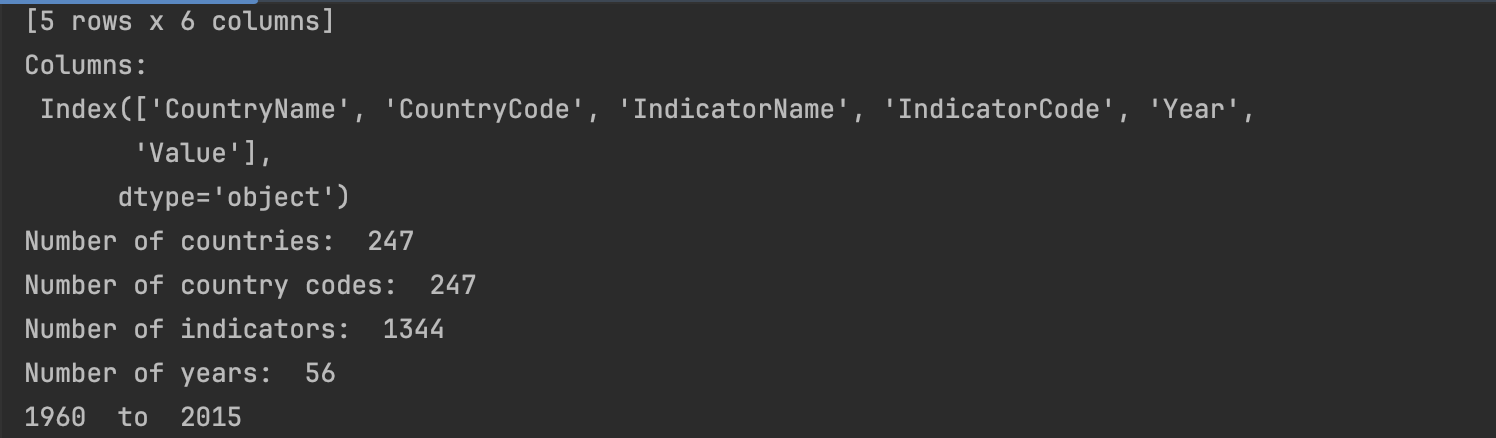
Reference : **8 ADVANCED MACHINE LEARNING PROJECTS:**

<https://drive.google.com/drive/folders/1Xe3HftLxL1T6HsEBUfjq_zXANjTnr6Cz?usp=sharing>

import pandas as pd  
import numpy as np  
import random  
import matplotlib.pyplot as plt  
import matplotlib.cbook  
import zipfile  
import bz2  
import warnings  
warnings.filterwarnings("ignore", category=matplotlib.MatplotlibDeprecationWarning)  
  
data = pd.read\_csv('data/Indicators.bz2')  
print("data.shape: ", data.shape)  
print("Sample Data: \n", data.head())  
print("Columns: \n", data.columns)

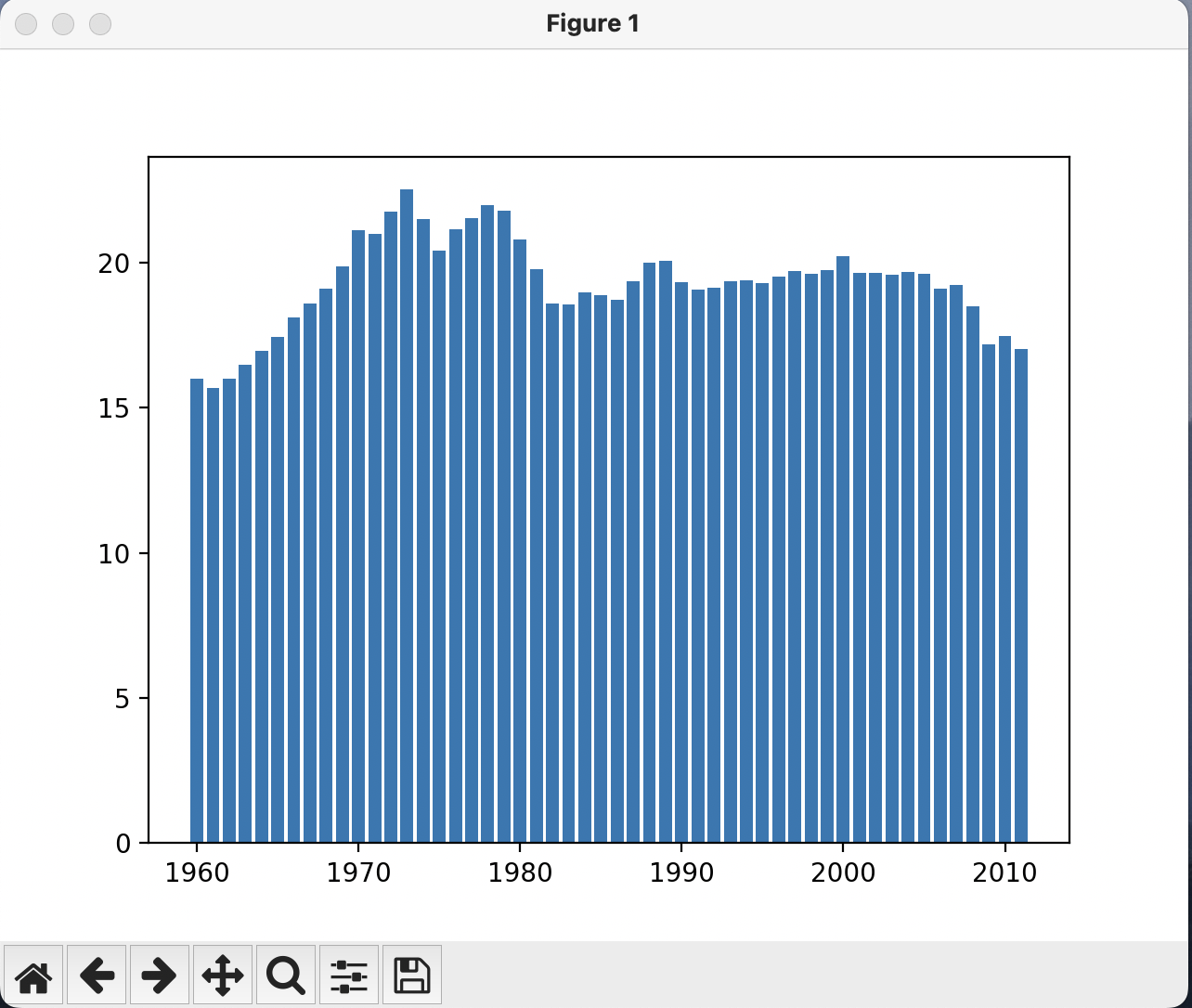


countries = data['CountryName'].unique().tolist()  
print("Number of countries: ", len(countries))  
  
countryCodes = data['CountryCode'].unique().tolist()  
print("Number of country codes: ", len(countryCodes))  
  
indicators = data['IndicatorName'].unique().tolist()  
print("Number of indicators: ", len(indicators))  
  
years = data['Year'].unique().tolist()  
print("Number of years: ", len(years))  
  
print(min(years)," to ", max(years))

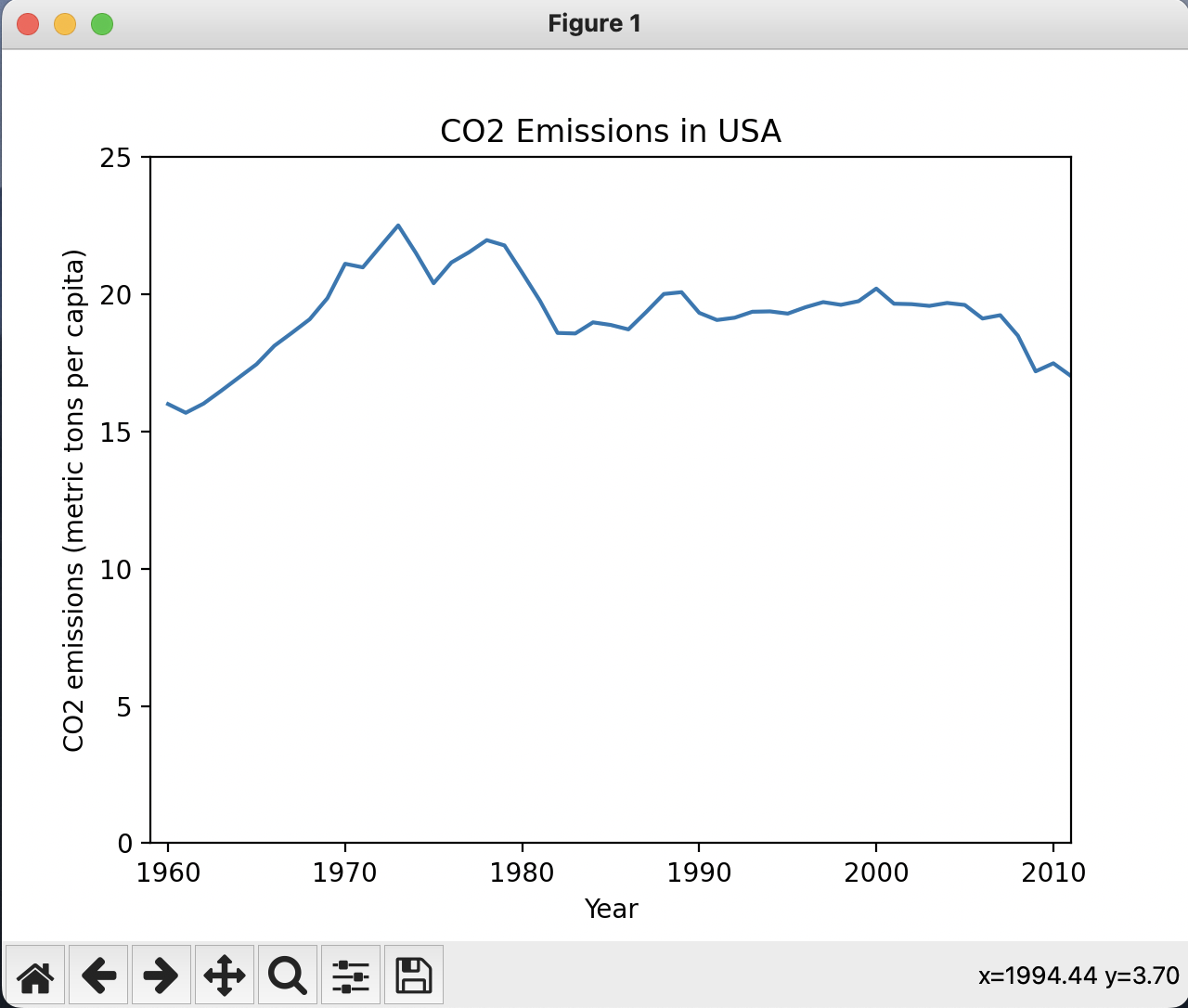


hist\_indicator = 'CO2 emissions \(metric'  
hist\_country = 'USA'  
mask1 = data['IndicatorName'].str.contains(hist\_indicator)  
mask2 = data['CountryCode'].str.contains(hist\_country)  
stage = data[mask1 & mask2]  
  
print(stage.shape)  
stage.head()  
print("Indicator Name: ", stage['IndicatorName'].iloc[0])

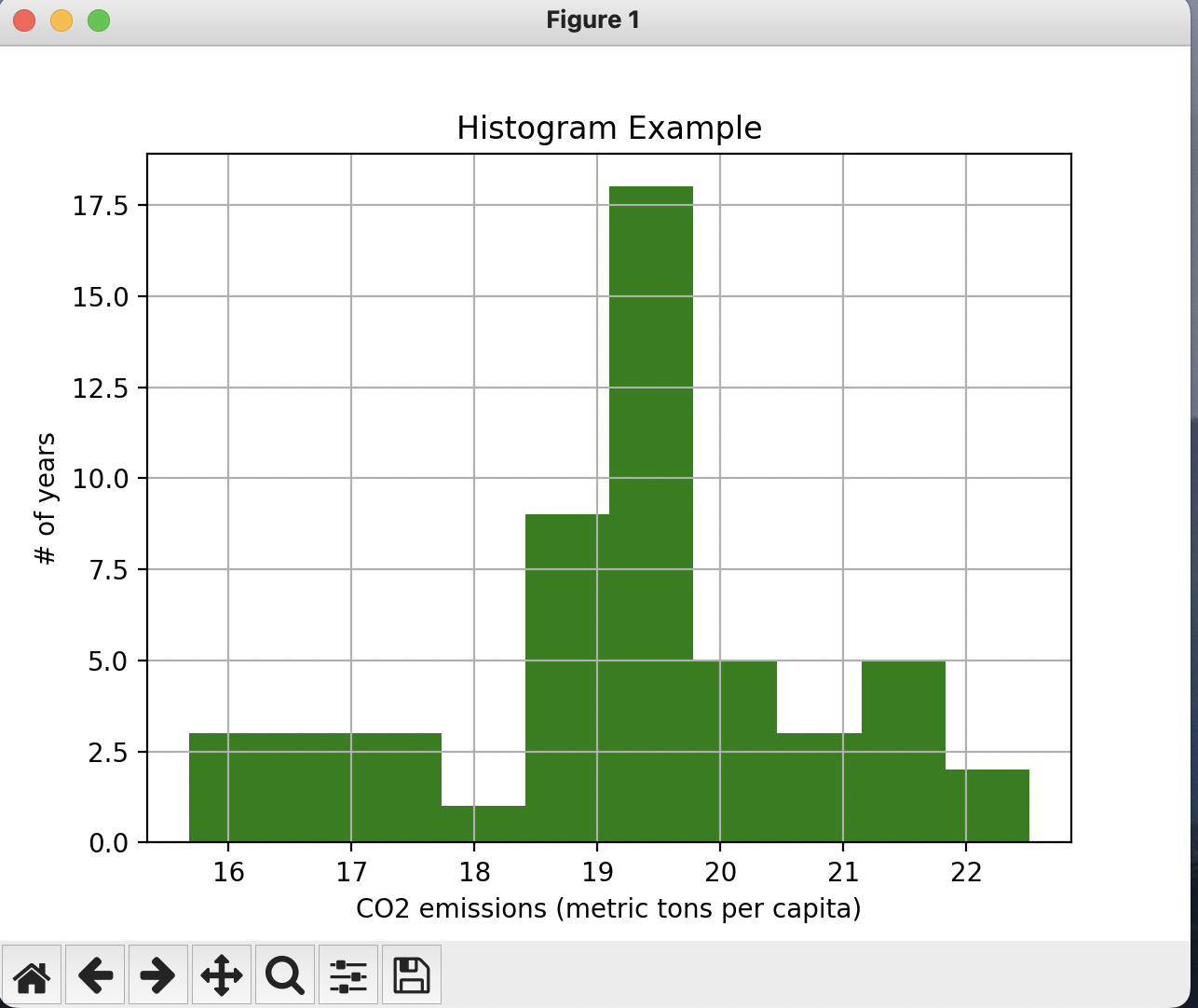
years = stage['Year'].values  
co2 = stage['Value'].values  
plt.bar(years, co2)  
plt.show()



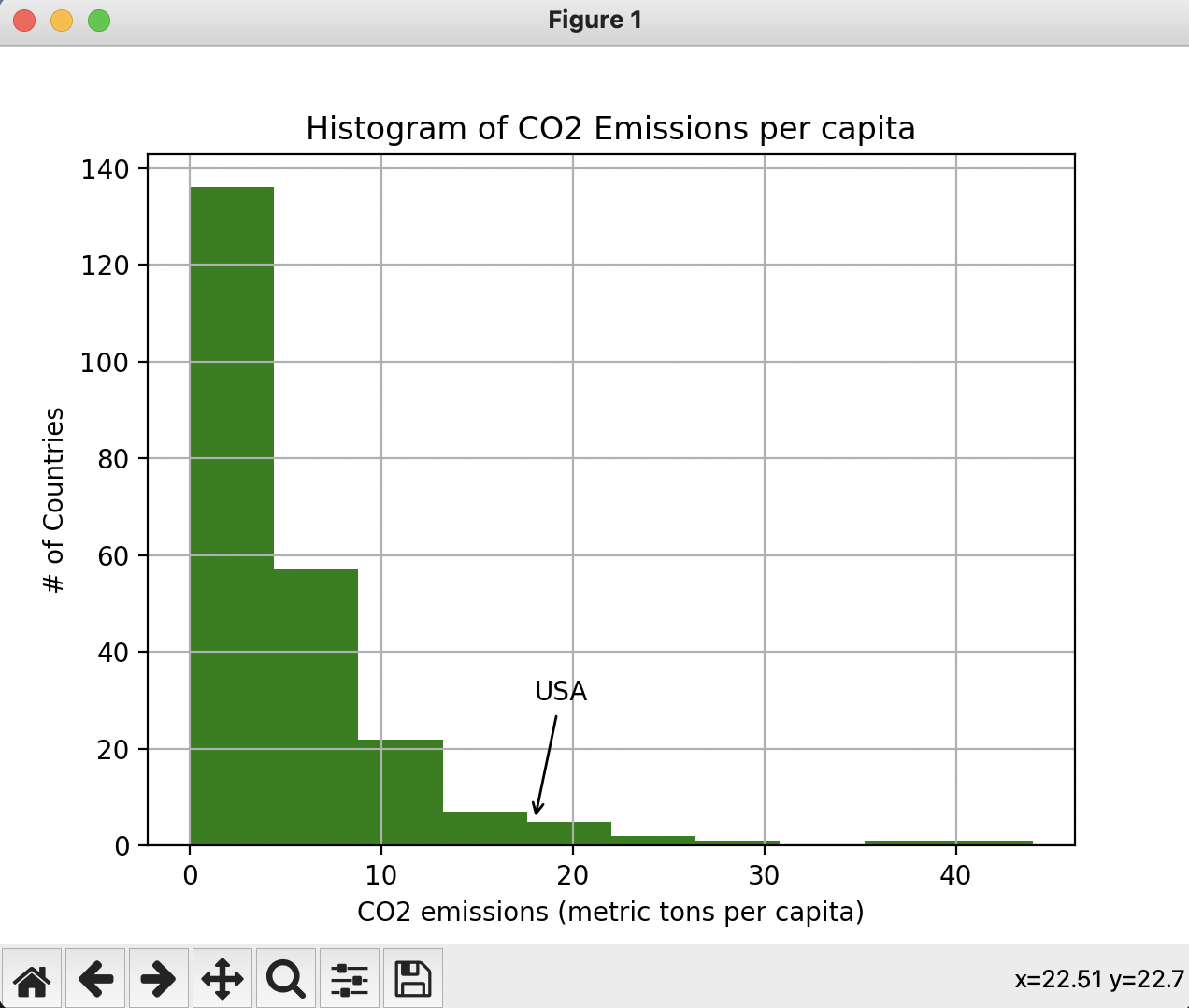
plt.plot(stage['Year'].values,stage['Value'].values)  
  
plt.xlabel('Year')  
plt.ylabel(stage['IndicatorName'].iloc[0])  
plt.title('CO2 Emissions in USA')  
plt.axis([1959, 2011, 0, 25])  
plt.show()



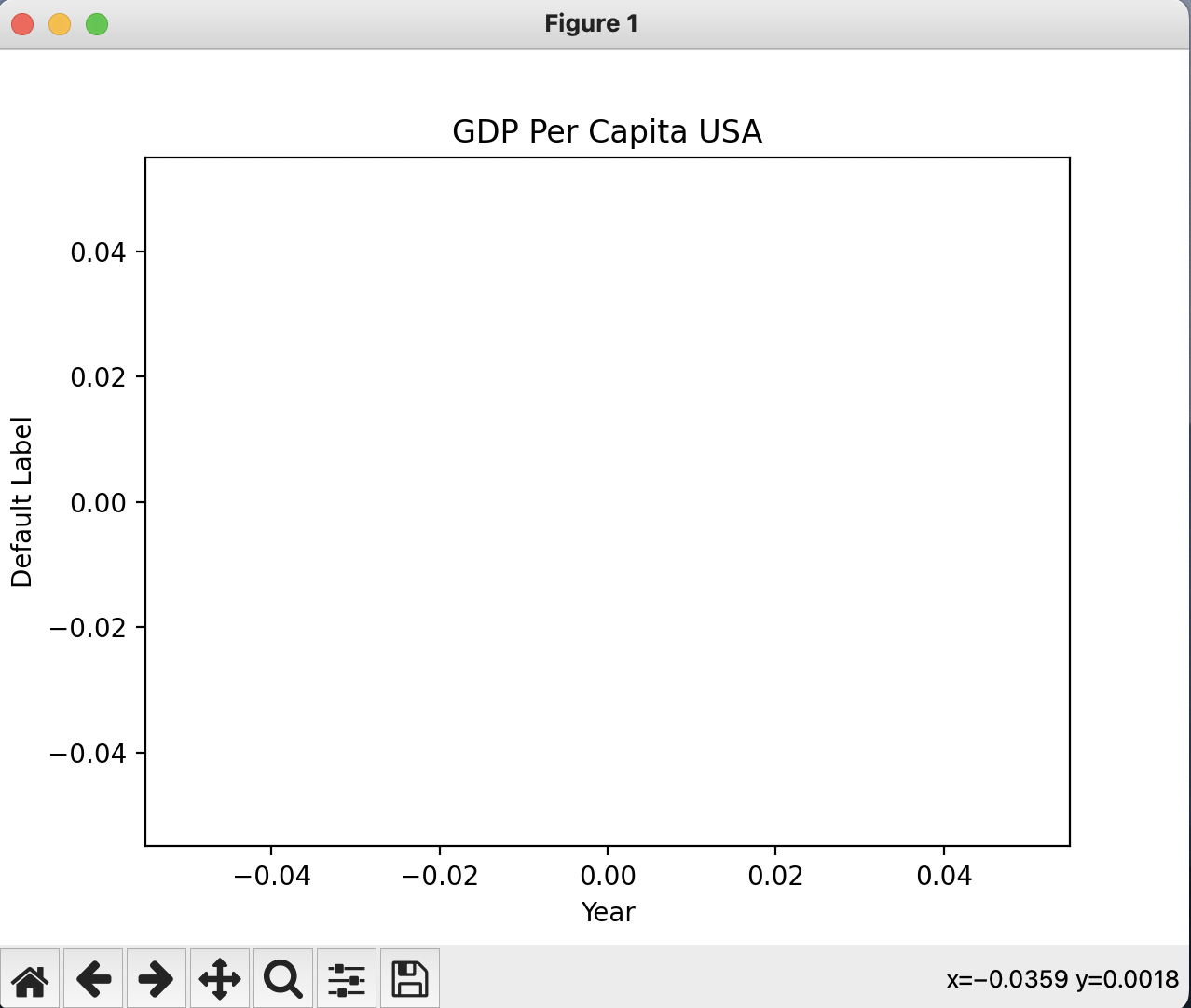
hist\_data = stage['Value'].values  
print(hist\_data)  
print(len(hist\_data))  
plt.hist(hist\_data, 10, density=False, facecolor='green')  
plt.xlabel(stage['IndicatorName'].iloc[0])  
plt.ylabel('# of years')  
plt.title('Histogram Example')  
plt.grid(True)  
plt.show()



hist\_indicator = 'CO2 emissions \(metric'  
hist\_year = 2011  
mask1 = data['IndicatorName'].str.contains(hist\_indicator)  
mask2 = data['Year'].isin([hist\_year])  
  
co2\_2011 = data[mask1 & mask2]  
co2\_2011.head()  
print(len(co2\_2011))  
fig, ax = plt.subplots()  
ax.annotate("USA", xy=(18, 5), xycoords='data', xytext=(18, 30), textcoords='data'  
 ,arrowprops=dict(arrowstyle="->", connectionstyle="arc3"))  
plt.hist(co2\_2011['Value'], 10, density=False, facecolor="green")  
plt.xlabel(stage['IndicatorName'].iloc[0])  
plt.ylabel('# of Countries')  
plt.title('Histogram of CO2 Emissions per capita')  
plt.grid(True)  
plt.show()



hist\_indicator = 'GDP per Capita \(constant 2005'  
hist\_country = 'USA'  
mask1 = data['IndicatorName'].str.contains(hist\_indicator)  
mask2 = data['CountryCode'].str.contains(hist\_country)  
  
gdp\_stage = data[mask1 & mask2]  
  
print("GDP: ",gdp\_stage.head())  
stage.head()  
  
plt.plot(gdp\_stage['Year'].values, gdp\_stage['Value'].values)  
  
plt.xlabel('Year')  
plt.ylabel(gdp\_stage['IndicatorName'].iloc[0])  
  
plt.title('GDP Per Capita USA')  
plt.show()  
  
print("GDP Min Year = ", gdp\_stage['Year'].min(), "max: ",  
 gdp\_stage['Year'].max())  
print("CO2 Min Year = ", stage['Year'].min(), "max: ", stage['Year'].max())  
  
gdp\_stage\_trunc = gdp\_stage[gdp\_stage['Year'] < 2012]  
print(len(gdp\_stage\_trunc))  
print(len(stage))  
  
import matplotlib.pyplot as plt  
fig, axis = plt.subplots()  
axis.yaxis.grid(True)  
axis.set\_title('CO2 Emissions vs. GDP (per capita)',fontsize=10)  
axis.set\_xlabel(gdp\_stage\_trunc['IndicatorName'].iloc[0],fontsize=10)  
axis.set\_ylabel(stage['IndicatorName'].iloc[0],fontsize=10)  
X = gdp\_stage\_trunc['Value']  
Y = stage['Value']  
axis.scatter(X, Y)  
plt.show()  
  
print(np.corrcoef(gdp\_stage\_trunc['Value'],stage['Value']))



Error screenshot :

