

▼ Chapter 6 - Exercise 2: Phân tích dữ liệu thế giới qua các năm

Cho các dữ liệu *year*, *pop*, *gdp_cap*, *life_exp*, *pop2*, *col* từ tập tin *data_year_pop_cap_life.txt*

```
# Năm thu thập dữ liệu
year = [1950, 1951, 1952, 1953, 1954, 1955, 1956, 1957, 1958, 1959, 1960, 1961, 1962, 1963]
# Dân số thế giới thực tế và dự đoán tương ứng với năm (year)
pop = [2.53, 2.57, 2.62, 2.67, 2.71, 2.76, 2.81, 2.86, 2.92, 2.97, 3.03, 3.08, 3.14, 3.2,

# Thu nhập bình quân đầu người và tuổi thọ trung bình của một số quốc gia
gdp_cap = [974.5803384, 5937.029525999998, 6223.367465, 4797.231267, 12779.37964, 34435.36
life_exp = [43.828, 76.423, 72.301, 42.731, 75.32, 81.235, 79.829, 75.635, 64.062, 79.441,

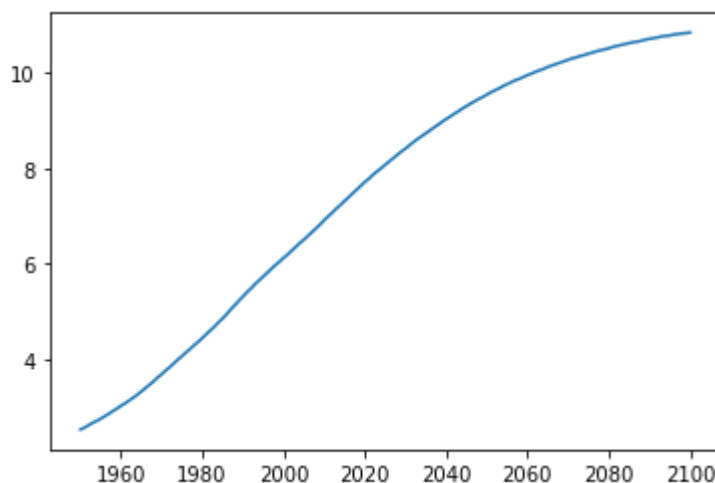
# life_exp1950 tương tự như life_exp nhưng của một số quốc gia khác
life_exp1950 = [28.8, 55.23, 43.08, 30.02, 62.48, 69.12, 66.8, 50.94, 37.48, 68.0, 38.22,

import matplotlib.pyplot as plt
import numpy as np

# Câu 1: In item cuối của year và pop
print(year[-1])
print(pop[-1])

    2100
    10.85

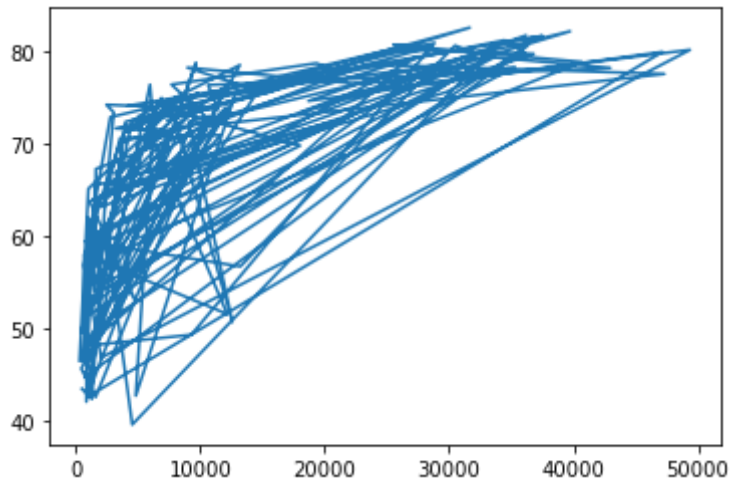
# Câu 2: Vẽ biểu đồ line thể hiện sự thay đổi dân số thế giới qua các năm ( x-axis: year,
plt.plot(year, pop)
plt.show()
```



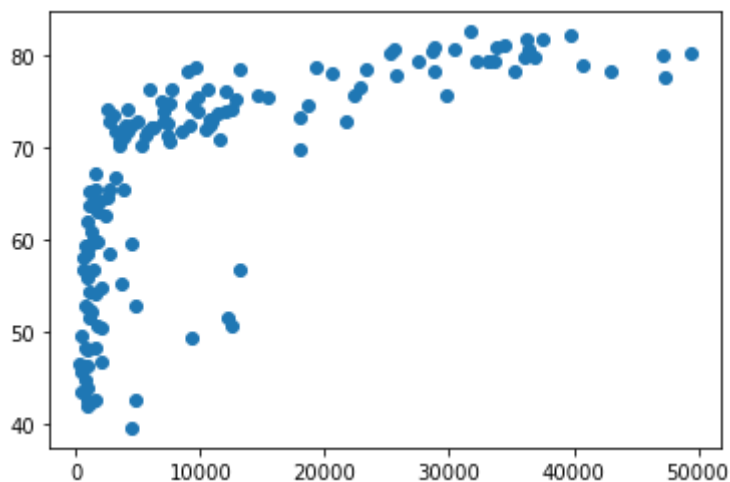
```
# Câu 3: Cho biết thu nhập bình quân đầu người và tuổi thọ trung bình của item cuối trong
print('Thu nhập bình quân đầu người của item cuối là:', round(gdp_cap[-1],2))
print('Tuổi thọ trung bình của item cuối là:', round(life_exp[-1],1))
```

Thu nhập bình quân đầu người của item cuối là: 469.71
Tuổi thọ trung bình của item cuối là: 43.5

```
# Câu 4: Thử vẽ biểu đồ line liên hệ giữa gdp_cap và life_exp với x-axis: gdp_cap, y-axis:
plt.plot(gdp_cap, life_exp)
plt.show()
# Biểu đồ này có thể xem được không? Nếu không thì bạn hãy đề xuất một loại biểu đồ phù hợp
```



```
# Có thể thay biểu đồ line thành biểu đồ scatter plot
plt.scatter(gdp_cap, life_exp)
plt.show()
```



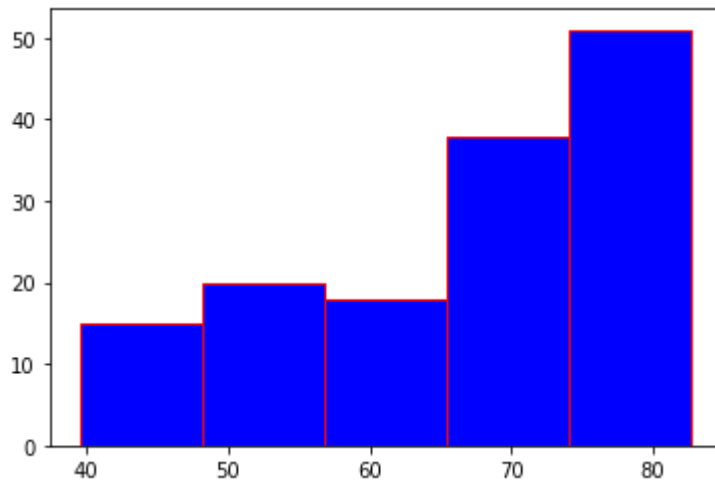
```
# Câu 5: Vẽ biểu đồ histogram của life_exp, màu cột xanh, viền đỏ
# Bạn nhận xét gì qua biểu đồ vừa vẽ
plt.hist(life_exp, color='g', edgecolor='r')
plt.show()
```



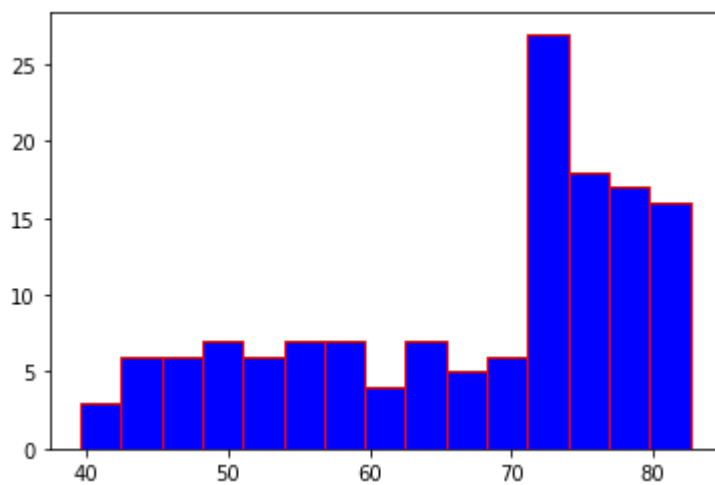
Câu 6: Vẽ biểu đồ histogram của life_exp, màu cột xanh dương, viền đỏ, với bins = 5, 15,
Bạn nhận xét gì qua các biểu đồ vừa vẽ ?



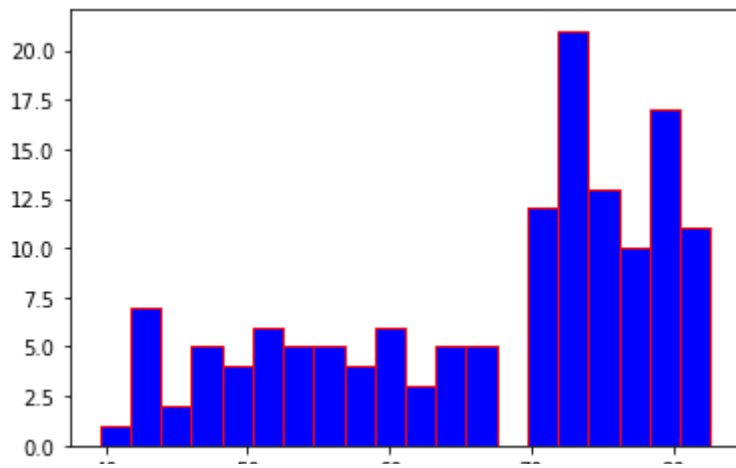
Với bins = 5
plt.hist(life_exp, bins=5, color='b', edgecolor='r')
plt.show()



Với bins = 15
plt.hist(life_exp, bins=15, color='b', edgecolor='r')
plt.show()



Với bins = 20
plt.hist(life_exp, bins=20, color='b', edgecolor='r')
plt.show()

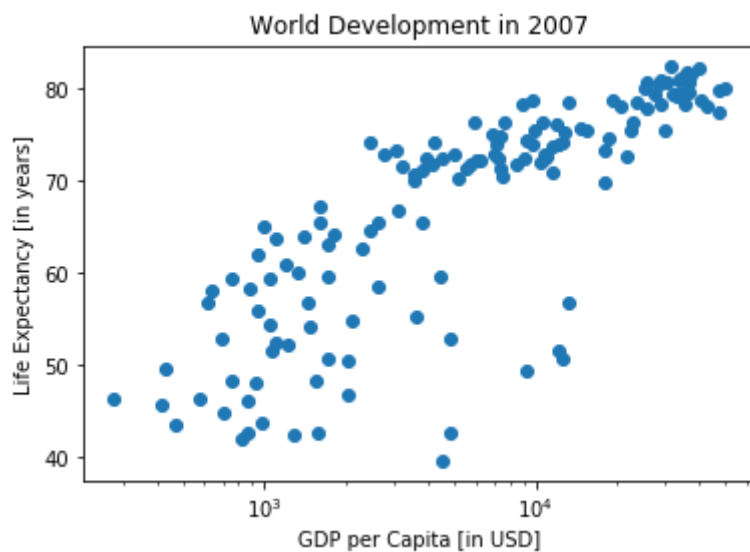


Câu 7: Tạo scatter plot của gdp_cap và life_exp nhưng sử dụng plt.xscale('log').
 # Khi trực quan hóa dữ liệu thay đổi trong phạm vi rất rộng, thang đo logarit plt.xscale('log')
 # chúng ta hình dung các thay đổi một cách trực quan hơn.

```
plt.scatter(gdp_cap, life_exp)
plt.xscale('log')

# Thiết lập xlabel, ylabel, title
plt.xlabel('GDP per Capita [in USD]')
plt.ylabel('Life Expectancy [in years]')
plt.title('World Development in 2007')

plt.show()
```



Câu 8: Tạo Scatter plot của gdp_cap và life_exp, sử dụng plt.xscale('log'). Thiết lập x1
 # Với: tick_val = [1000,10000,100000] và tick_lab = ['1k','10k','100k'] => plt.xticks(tick_val, tick_lab)

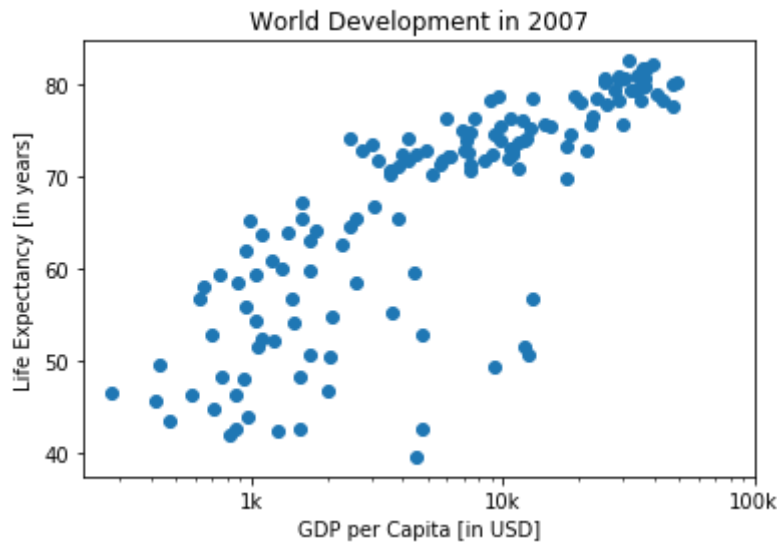
```
plt.scatter(gdp_cap, life_exp)

plt.xscale('log')
plt.xlabel('GDP per Capita [in USD]')
plt.ylabel('Life Expectancy [in years]')
plt.title('World Development in 2007')

tick_val = [1000,10000,100000]
tick_lab = ['1k','10k','100k']
```

```
tick_val = [ 1k , 10k , 100k ]
plt.xticks(tick_val, tick_lab)
```

```
plt.show()
```



Câu 9: Tạo numpy array np_pop từ pop2 trong file dữ liệu:

```
pop2 = [31.889923, 3.600523, 33.333216, 12.420476, 40.301927, 20.434176, 8.199783, 0.70857]
```

```
np_pop = np.array(pop2)
```

Vẽ scatter plot của gdp_cap và life_exp, với s = np_pop * 2, màu magenta

```
plt.scatter(gdp_cap, life_exp , s = np_pop * 2, color='m')
```

Thiết lập xlabel, ylabel, title và plt.xticks([1000, 10000, 100000],['1k', '10k', '100k'])

```
plt.xscale('log')
```

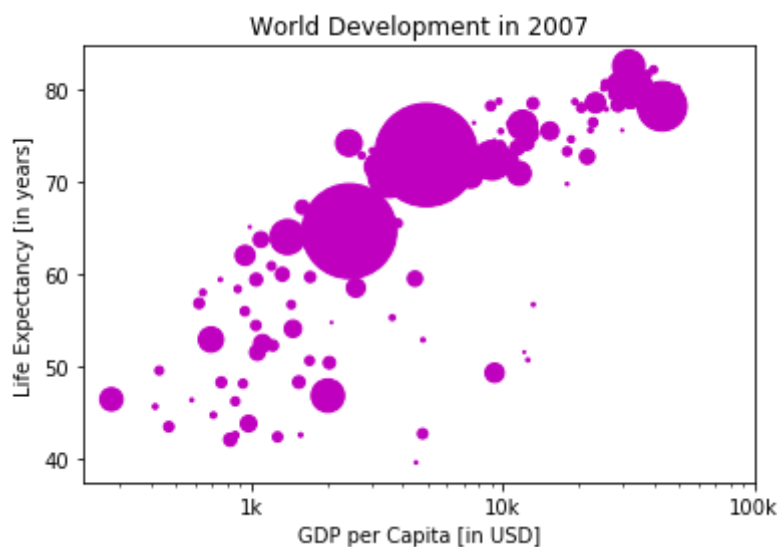
```
plt.xlabel('GDP per Capita [in USD]')
```

```
plt.ylabel('Life Expectancy [in years]')
```

```
plt.title('World Development in 2007')
```

```
plt.xticks([1000, 10000, 100000],['1k', '10k', '100k'])
```

```
plt.show()
```



```
# Câu 10:
# Danh sách màu tương ứng với các nhóm quốc gia
# 'Asia': 'red', 'Europe': 'green', 'Africa': 'blue', 'Americas': 'yellow', 'Oceania': 'black'

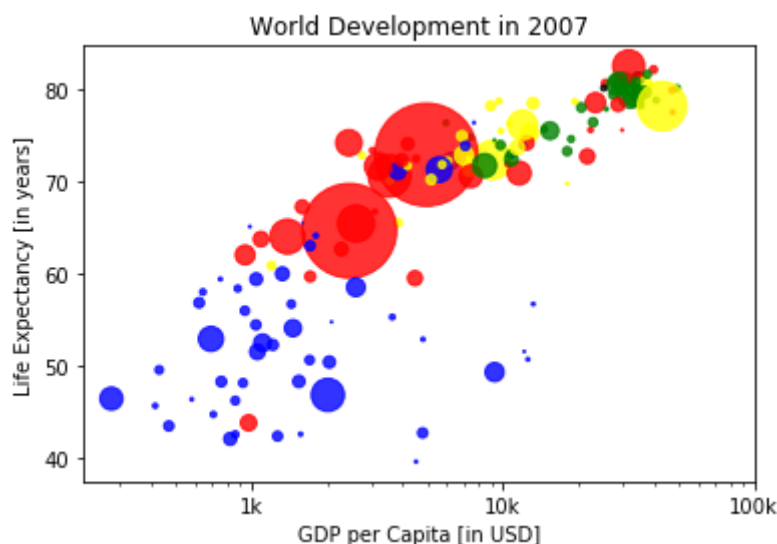
col = ['red', 'green', 'blue', 'blue', 'yellow', 'black', 'green', 'red', 'red', 'green',
       'yellow', 'green', 'blue', 'yellow', 'green', 'blue', 'blue', 'red', 'blue', 'yellow',
       'blue', 'yellow', 'red', 'yellow', 'blue', 'blue', 'blue', 'yellow', 'blue', 'green',
       'green', 'blue', 'yellow', 'yellow', 'blue', 'yellow', 'blue', 'blue', 'blue', 'green',
       'blue', 'green', 'blue', 'green', 'yellow', 'blue', 'blue', 'yellow', 'yellow', 'red',
       'red', 'red', 'red', 'red', 'green', 'red', 'green', 'yellow', 'red', 'red', 'blue',
       'red', 'blue', 'blue', 'blue', 'blue', 'blue', 'red', 'blue', 'blue', 'blue', 'yellow',
       'blue', 'blue', 'red', 'blue', 'red', 'green', 'black', 'yellow', 'blue', 'blue', 'yellow',
       'yellow', 'yellow', 'yellow', 'red', 'green', 'green', 'yellow', 'blue', 'green', 'blue',
       'green', 'blue', 'red', 'green', 'green', 'blue', 'blue', 'green', 'red', 'green',
       'red', 'red', 'blue', 'red', 'blue', 'yellow', 'blue', 'green', 'blue', 'green',
       'yellow', 'red', 'red', 'red', 'blue', 'blue']

# Vẽ scatter plot của gdp_cap và life_exp, với s = np.array(pop) * 2, màu c = col (giá trị
plt.scatter(x = gdp_cap, y = life_exp, s = np.array(pop2) * 2, c= col, alpha=0.8)

# Thiết lập xlabel, ylabel, title và plt.xticks([1000, 10000, 100000], ['1k', '10k', '100k'])
plt.xscale('log')
plt.xlabel('GDP per Capita [in USD]')
plt.ylabel('Life Expectancy [in years]')
plt.title('World Development in 2007')
plt.xticks([1000, 10000, 100000], ['1k', '10k', '100k'])

plt.show()

# Bạn nhận xét gì về biểu đồ vừa vẽ
```



```
# Câu 11: Vẽ scatter plot của gdp_cap, life_exp, với s = np.array(pop) * 2, màu c = col, a
plt.scatter(x = gdp_cap, y = life_exp, s = np.array(pop2) * 2, c = col, alpha = 0.8)

# Thiết lập xlabel, ylabel, title và plt.xticks([1000, 10000, 100000], ['1k', '10k', '100k'])
plt.xscale('log')
plt.xlabel('GDP per Capita [in USD]')
plt.ylabel('Life Expectancy [in years]')
```

```
plt.title('World Development in 2007')
plt.xticks([1000,10000,100000], ['1k','10k','100k'])

# Thêm text cho 2 nơi là India và China: plt.text(1550, 71, 'India'), plt.text(5700, 80, '
plt.text(1550, 71, 'India')
plt.text(5700, 80, 'China')

# Thêm lưới cho biểu đồ
plt.grid()

plt.show()
```

