

# BÀI TẬP THỰC HÀNH

## Bài 8:

# Trực quan hóa dữ liệu với Matplotlib

AI Academy Vietnam



# Thực hành 1

**Tập dữ liệu gas\_prices.csv:** Lưu trữ giá Gas của 10 nước trên thế giới trong giai đoạn từ năm 1990 - 2008

How to create a bar chart in

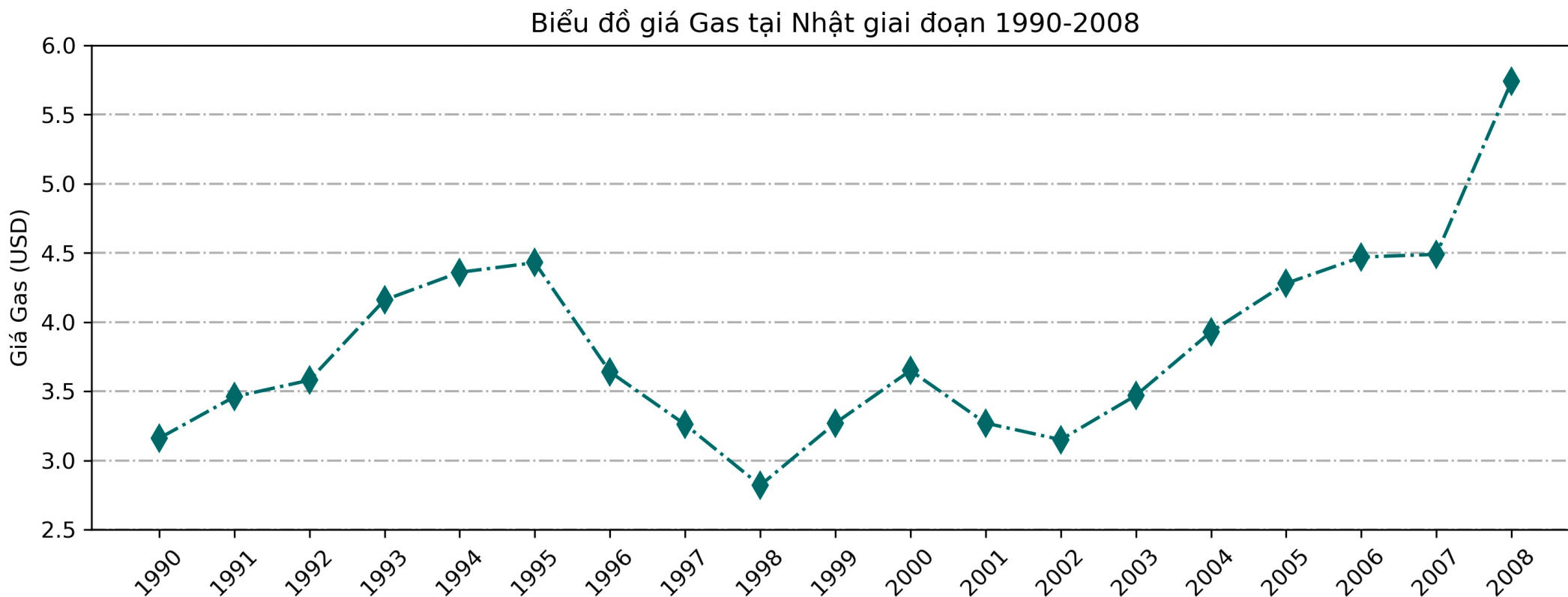
**Python with Matplotlib**



```
1 data = pd.read_csv('Data_Visualize/gas_prices.csv')
2 data
```

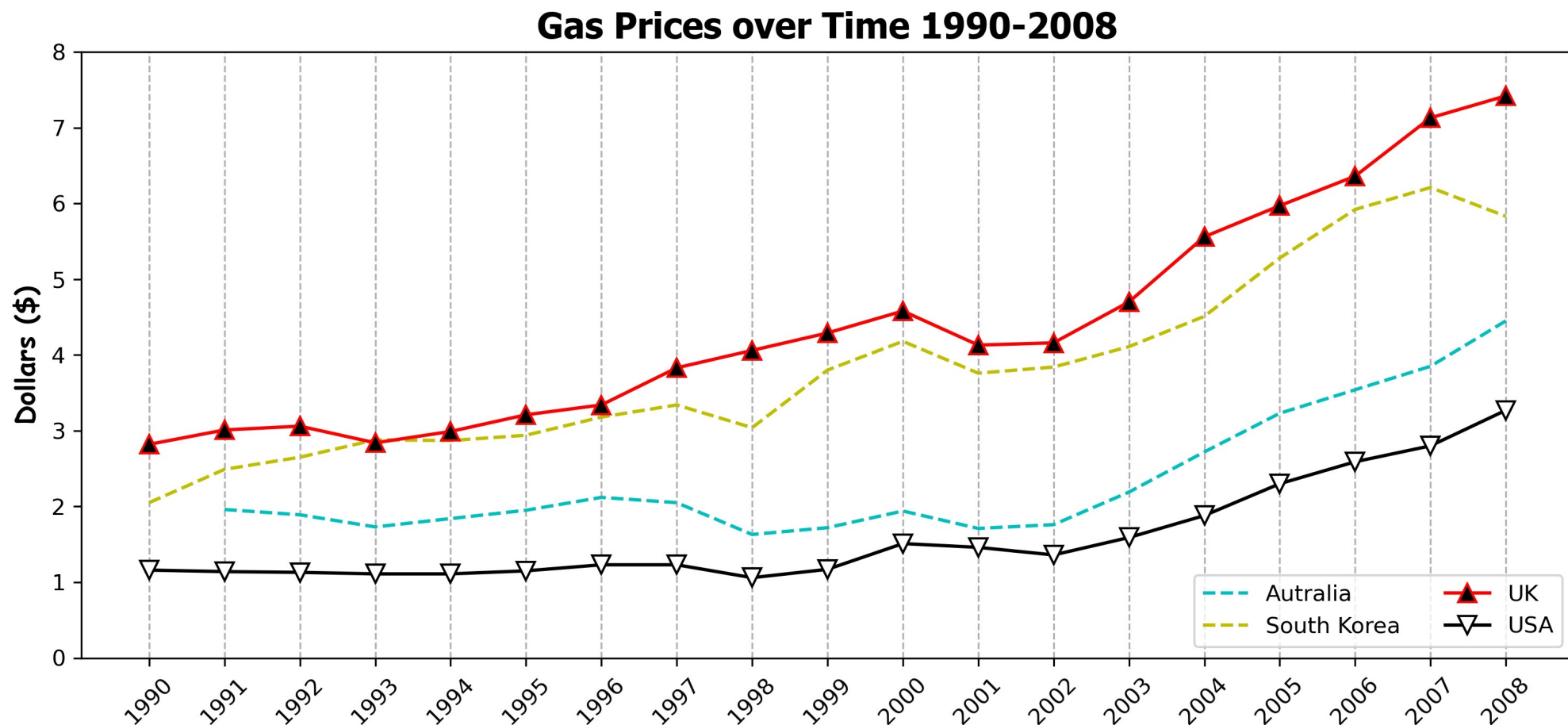
	Year	Australia	Canada	France	Germany	Italy	Japan	Mexico	South Korea	UK	USA
0	1990	NaN	1.87	3.63	2.65	4.59	3.16	1.00	2.05	2.82	1.16
1	1991	1.96	1.92	3.45	2.90	4.50	3.46	1.30	2.49	3.01	1.14
2	1992	1.89	1.73	3.56	3.27	4.53	3.58	1.50	2.65	3.06	1.13
3	1993	1.73	1.57	3.41	3.07	3.68	4.16	1.56	2.88	2.84	1.11
4	1994	1.84	1.45	3.59	3.52	3.70	4.36	1.48	2.87	2.99	1.11
5	1995	1.95	1.53	4.26	3.96	4.00	4.43	1.11	2.94	3.21	1.15
6	1996	2.12	1.61	4.41	3.94	4.39	3.64	1.25	3.18	3.34	1.23
7	1997	2.05	1.62	4.00	3.53	4.07	3.26	1.47	3.34	3.83	1.23
8	1998	1.63	1.38	3.87	3.34	3.84	2.82	1.49	3.04	4.06	1.06
9	1999	1.72	1.52	3.85	3.42	3.87	3.27	1.79	3.80	4.29	1.17
10	2000	1.94	1.86	3.80	3.45	3.77	3.65	2.01	4.18	4.58	1.51

**YÊU CẦU:** Vẽ biểu đồ giá gas của Japan, thiết lập các tham số để thu được biểu đồ như sau:



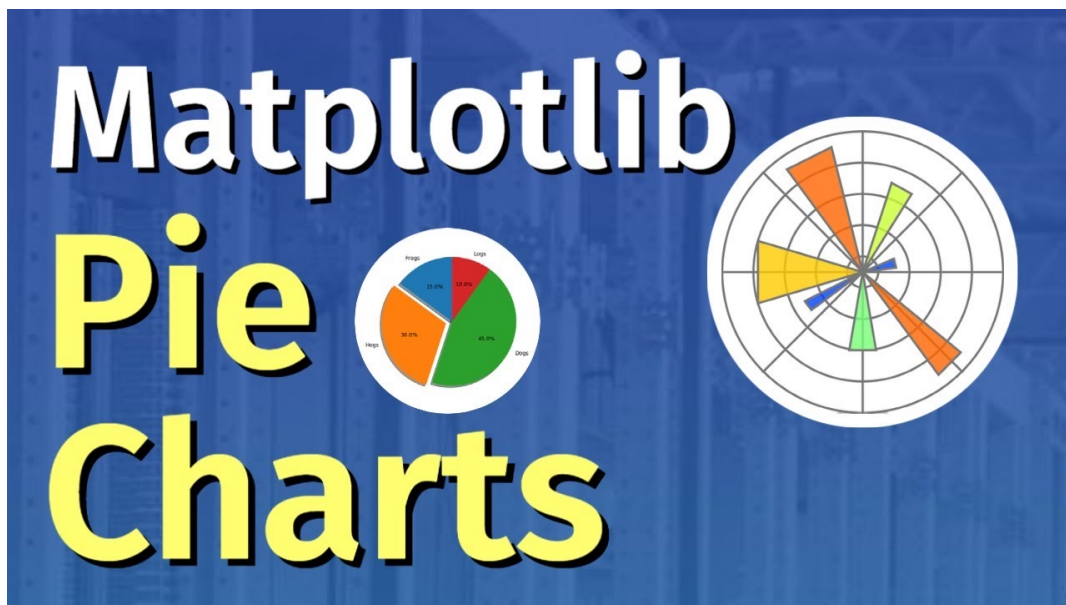
# Thực hành 2

**YÊU CẦU:** Trực quan hoá dữ liệu giá Gas của 4 nước: UK, USA, Australia, South Korea. với mục tiêu nhấn mạnh quốc gia có giá Gas cao nhất và thấp nhất. (như minh hoạ)



# Thực hành 3

**Số liệu:** Số lượng SV của các Khoa theo Giới tính.



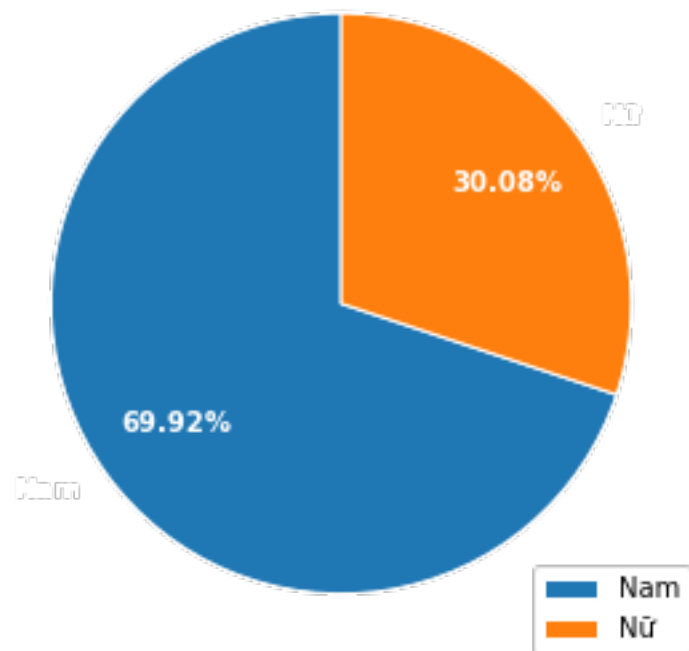
Khoá	K60	K61	K62	K63	K64	K65
Nam	200	340	260	440	300	180
Nữ	30	60	90	160	180	220

```
1  #Tạo dữ liệu: Số lượng SV Nam - Nữ theo từng khoá.
2  labels = ['K60', 'K61', 'K62', 'K63', 'K64', 'K65']
3  boys   = [200, 340, 260, 440, 300, 180]
4  girls  = [30, 60, 90, 160, 180, 220]
5  total  = list(np.array(boys) + np.array(girls))
6
7  sex    = ['Nam', 'Nữ']
8  sum_boy = sum(boys)
9  sum_girl = sum(girls)
```

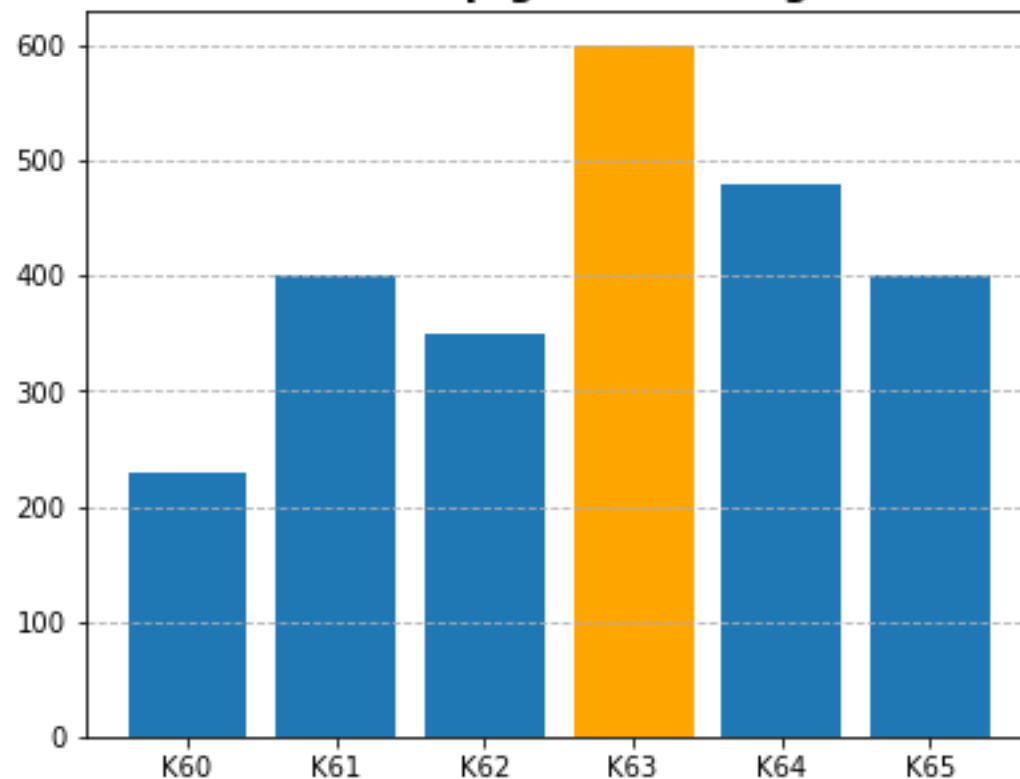


**Yêu cầu: Sử dụng số liệu đã có, thực hiện vẽ biểu đồ như minh họa dưới đây:**

**Biểu đồ tỷ lệ SV Nam - Nữ của Khoa CNTT**

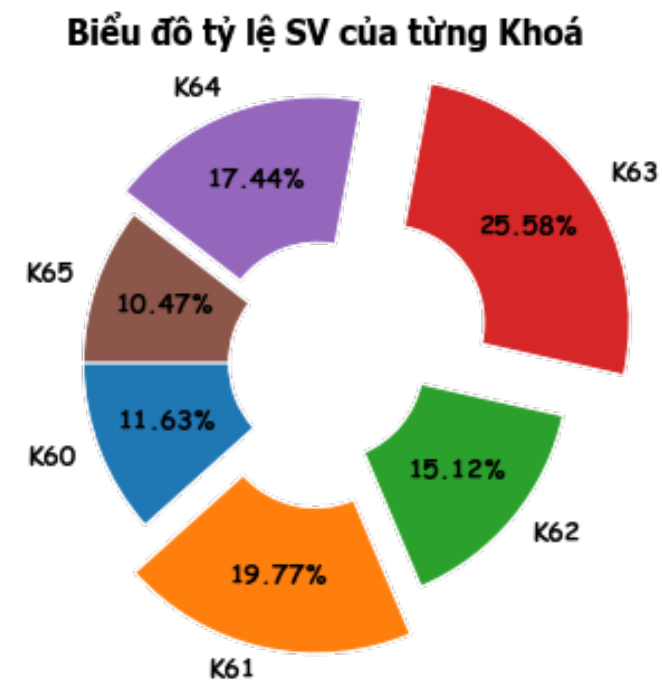
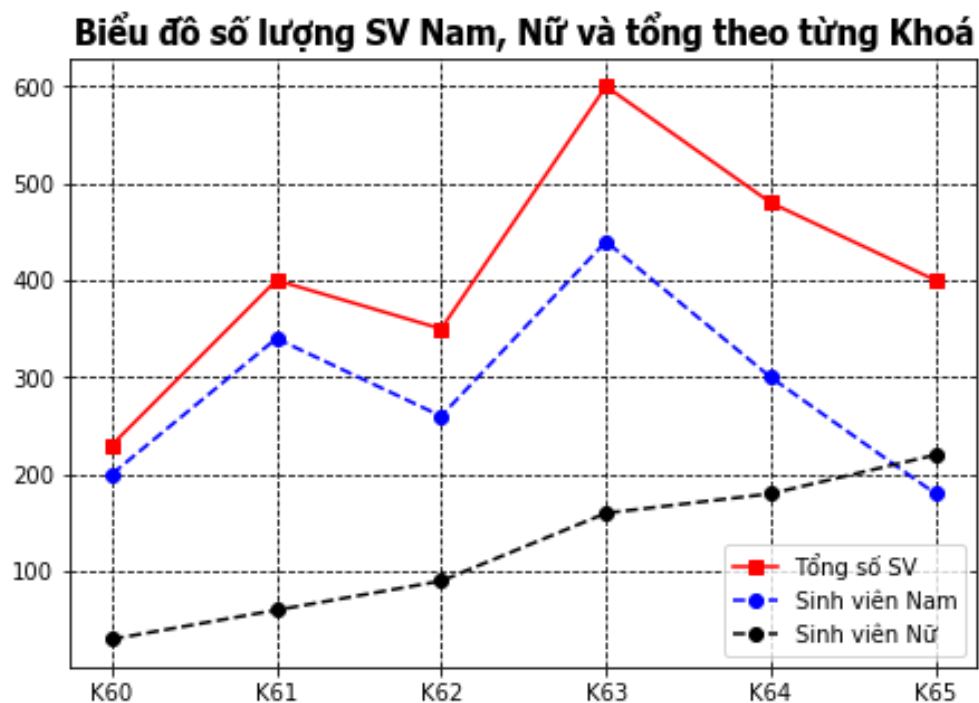


**Biểu đồ số lượng SV theo từng khoá**



# Thực hành 4

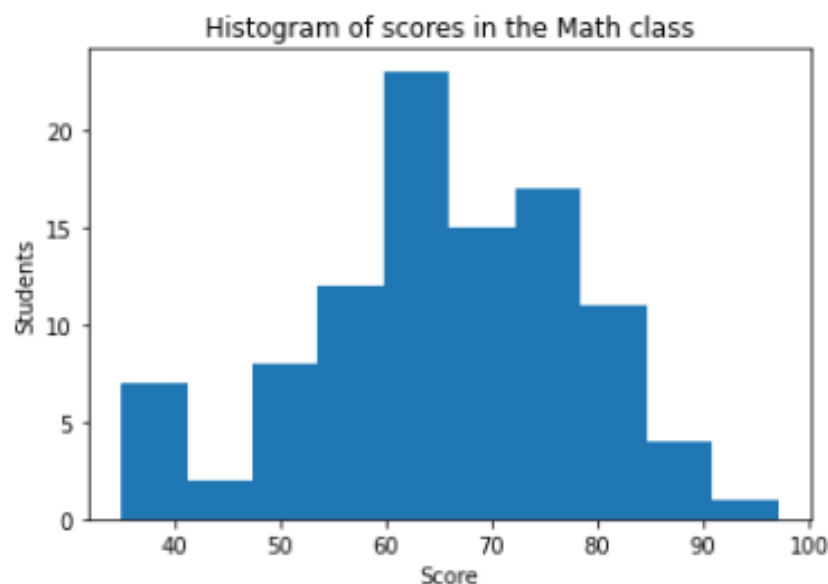
**Yêu cầu: Sử dụng số liệu đã có, thực hiện vẽ biểu đồ line chart và pie chart như minh họa dưới đây:**



# Thực hành 5

Tập dữ liệu **Data\_score.csv** lưu trữ dữ liệu 3 môn Math, Science, History của 500 học sinh

matplotlib



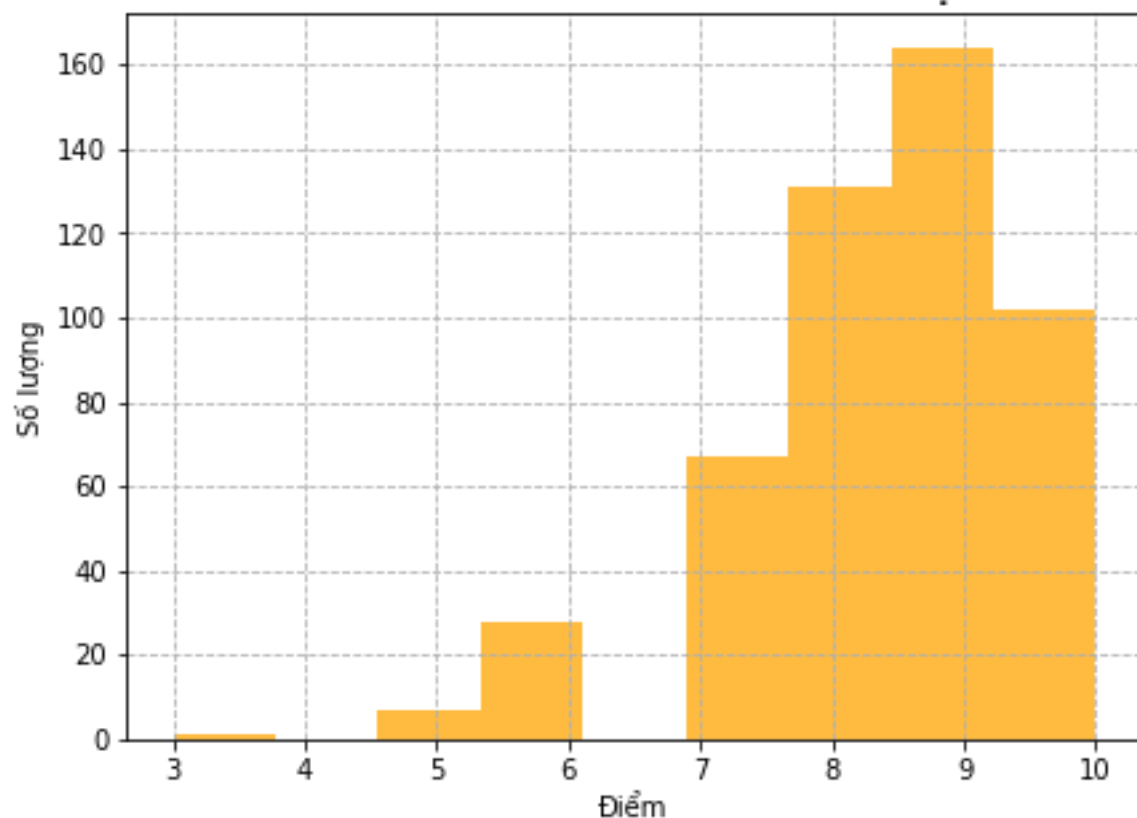
Plot a Histogram

	A	B	C
1	Math	Science	History
2	7	10	4
3	4	9	3
4	4	8	2
5	4	8	3
6	4	6	3
7	6	8	1
8	5	10	7
9	5	10	5
10	5	7	4
11	6	8	3
12	7	9	3
13	6	8	2
14	5	10	7
15	5	9	6
16	7	5	2
17	8	10	3
18	4	9	7
19	5	7	2
20	5	9	6

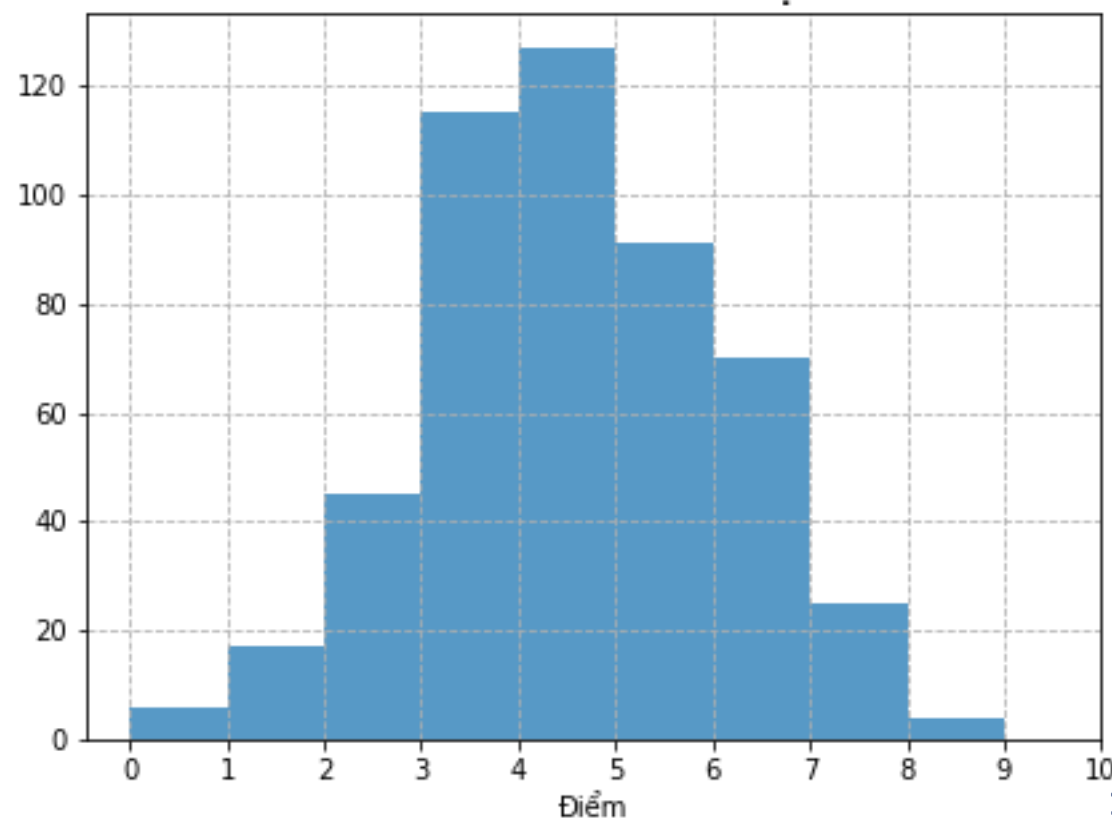
Sheet1 +

**YÊU CẦU:** Sử dụng các biểu đồ Histogram áp dụng cho môn Science và History --> nhận xét về điểm của 2 môn này

BIỂU ĐỒ PHỎ ĐỀM MÔN KHOA HỌC



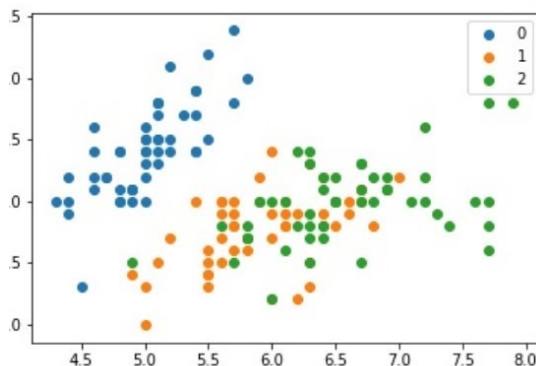
BIỂU ĐỒ PHỎ ĐỀM MÔN LỊCH SỬ



# Thực hành 6

Tập dữ liệu **Diamonds.txt** lưu trữ trọng lượng (cara) và giá (\$) tương ứng của 50 viên kim cương.

matplotlib



## MATPLOTLIB SCATTER PLOTS

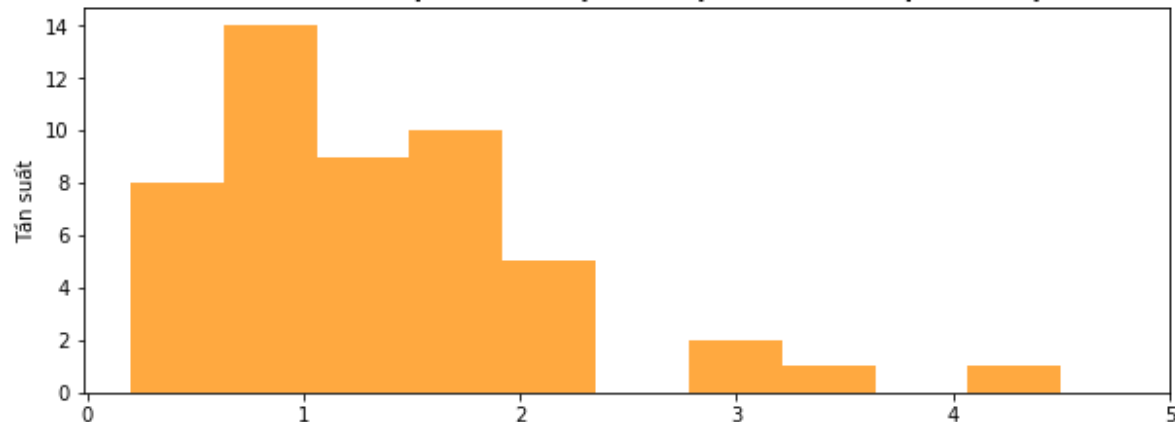
DATA VISUALIZATION PYTHON

Dia...	
0.23	484
0.31	942
0.2	345
1.02	4459
1.63	14022
1.14	4212
2.01	11925
1.28	9548
1.7	11605
1.01	4642
0.64	3541
0.97	4504
1.78	13691
3.4	15964
3.01	10453
1.51	11560
1.37	7979
1.5	9533
0.54	1723
0.72	3344
1.13	6133
2.24	13827
3.01	16538
4.5	18531
0.92	3625
1.05	7879
0.55	1319
0.74	2761
0.91	3620
1.23	6165

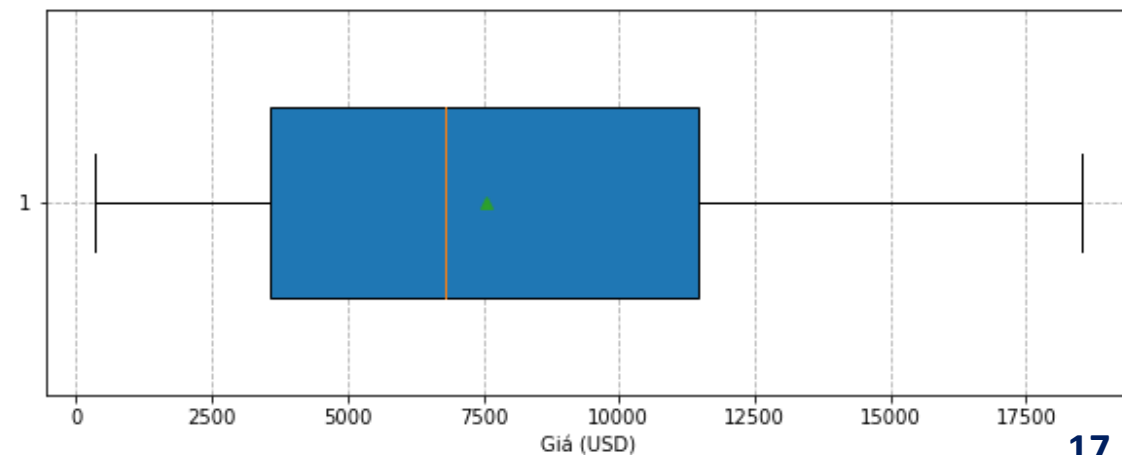
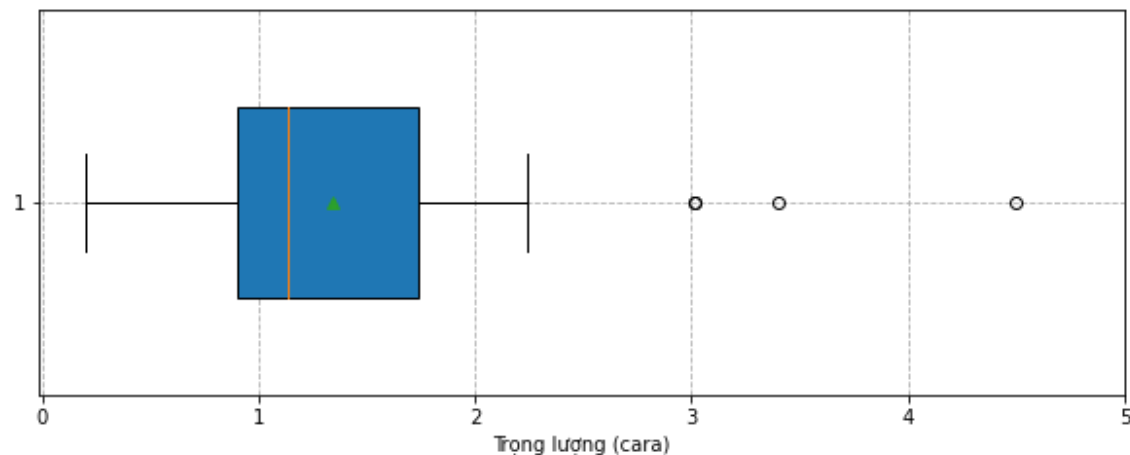
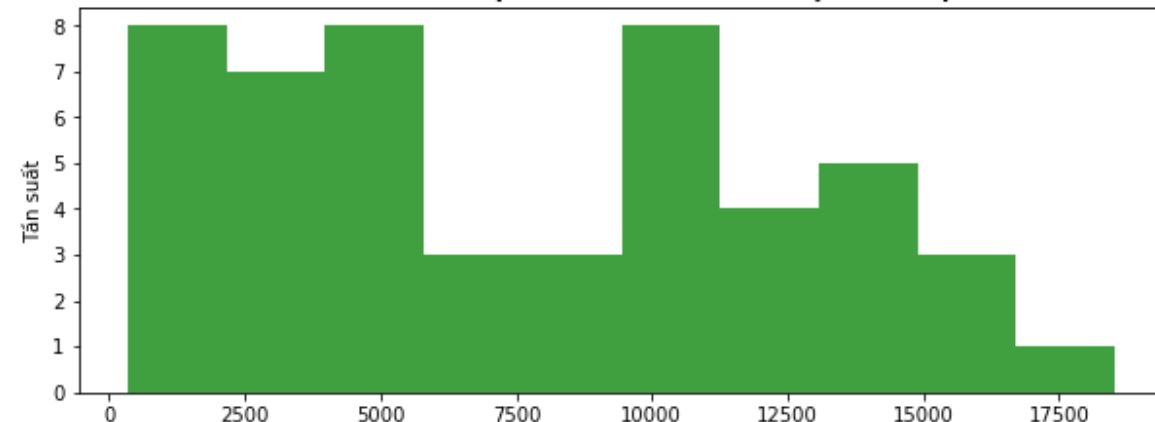


**YÊU CẦU:** Đọc dữ liệu file Diamonds.txt, sử dụng biểu đồ Boxplot cho dữ liệu trọng lượng và giá: Xác định các thông số Minimum, Maximum, Q1, Median, Mean, Q3, ngoại lai (nếu có) --> cho nhận xét từ biểu đồ thu được.

QUAN SÁT THUỘC TÍNH TRỌNG LƯỢNG TRONG TẬP DỮ LIỆU



QUAN SÁT THUỘC TÍNH GIÁ TRONG TẬP DỮ LIỆU



# Bài thực hành tổng hợp

## Mô tả tập dữ liệu Data\_500.csv:

Tập dữ liệu chứa thông tin của 500 người, bao gồm:

- **Gender** : Male / Female
- **Age** : Number (Age)
- **Height** : Number (cm)
- **Weight** : Number (Kg)
- **Index** :
  - 0 - Extremely Weak
  - 1 - Weak
  - 2 - Normal
  - 3 - Overweight
  - 4 - Obesity
  - 5 - Extreme Obesity

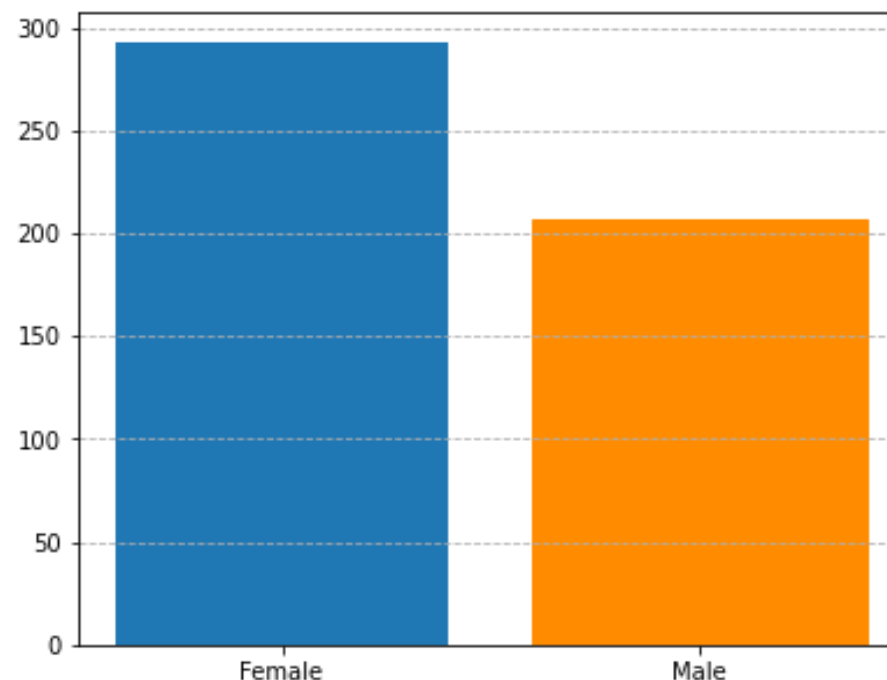
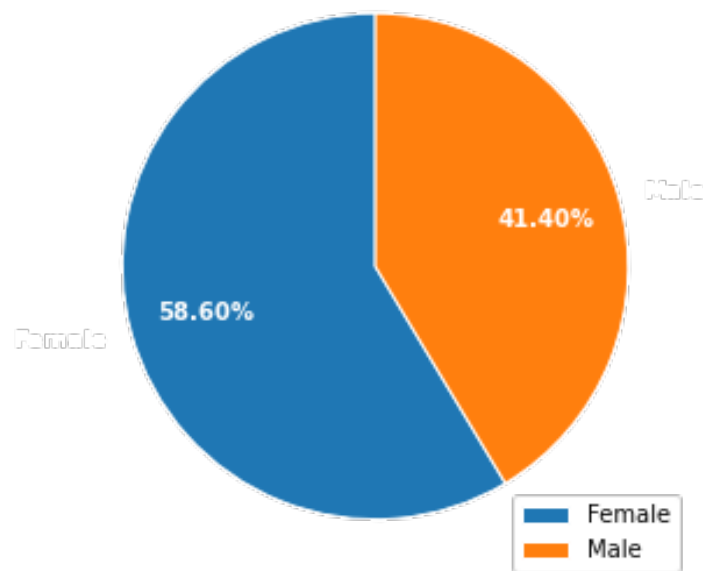
Data\_500

Gender	Age	Height	Weight	Index
Male	28	174	96	4
Male	24	189	87	2
Female	39	185	110	4
Female	49	195	104	3
Male	26	149	61	3
Male	26	189	104	3
Male	39	147	92	5

## Yêu cầu 1:

Thực hiện thống kê số lượng Nam – Nữ trong tập dữ liệu và trực quan hóa kết quả:

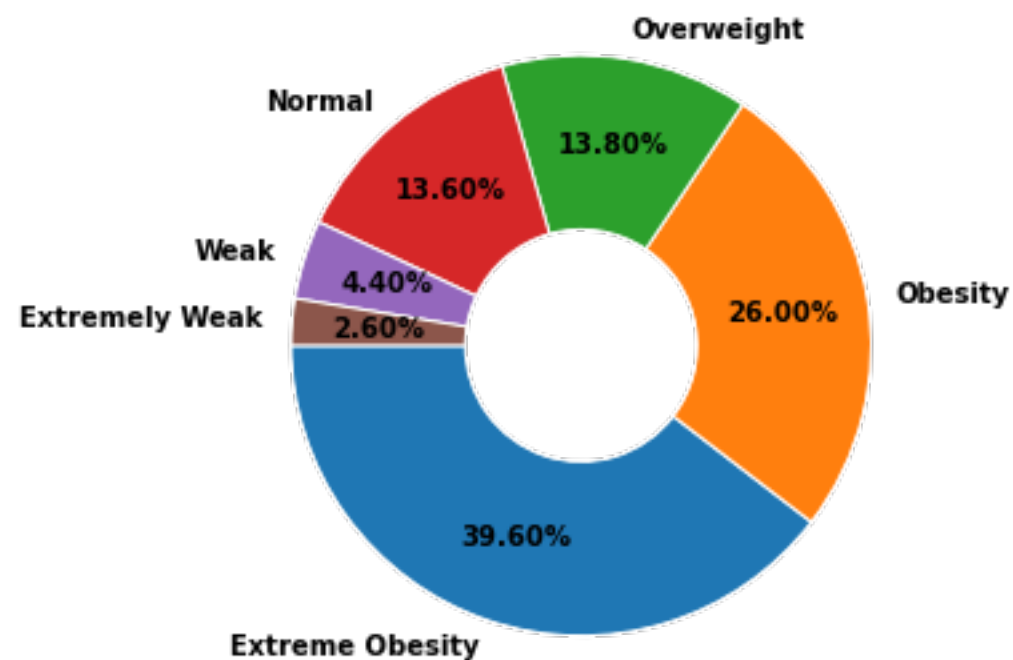
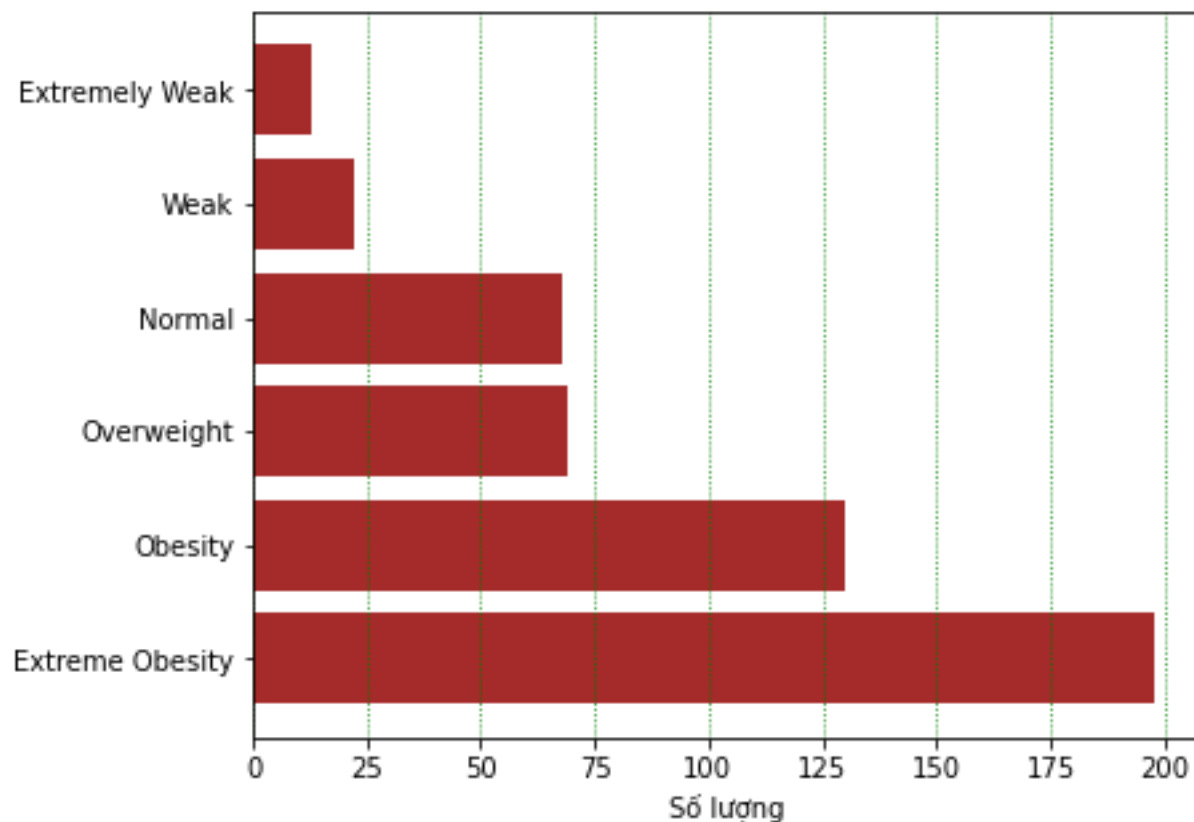
THỐNG KÊ DỮ LIỆU THEO GIỚI TÍNH TRONG TẬP DỮ LIỆU



## Yêu cầu 2:

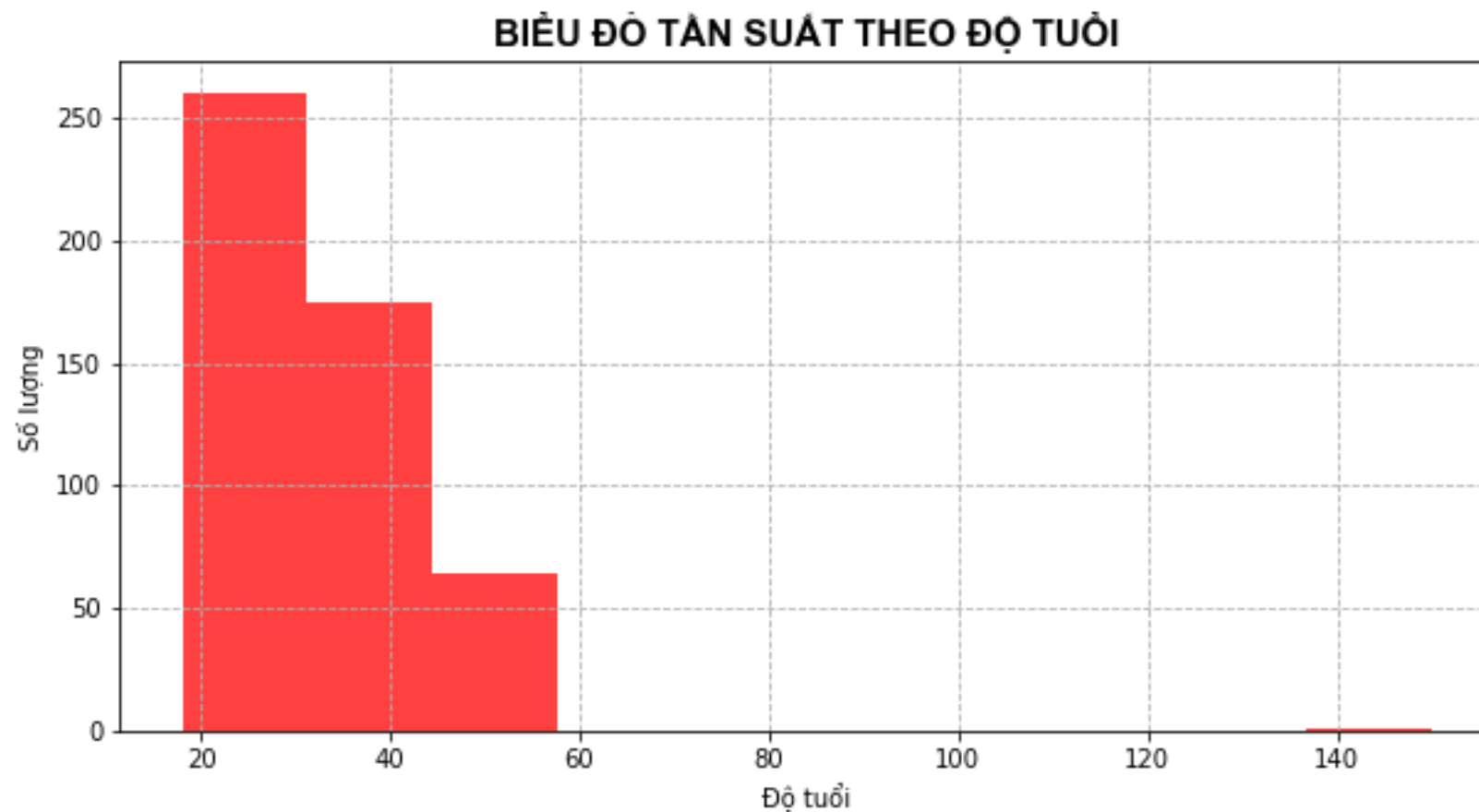
Thực hiện thống kê số lượng theo chỉ số cơ thể (index) và trực quan hóa kết quả.

**THỐNG KÊ DỮ LIỆU THEO TRẠNG THÁI CƠ THỂ TRONG TẬP DỮ LIỆU**



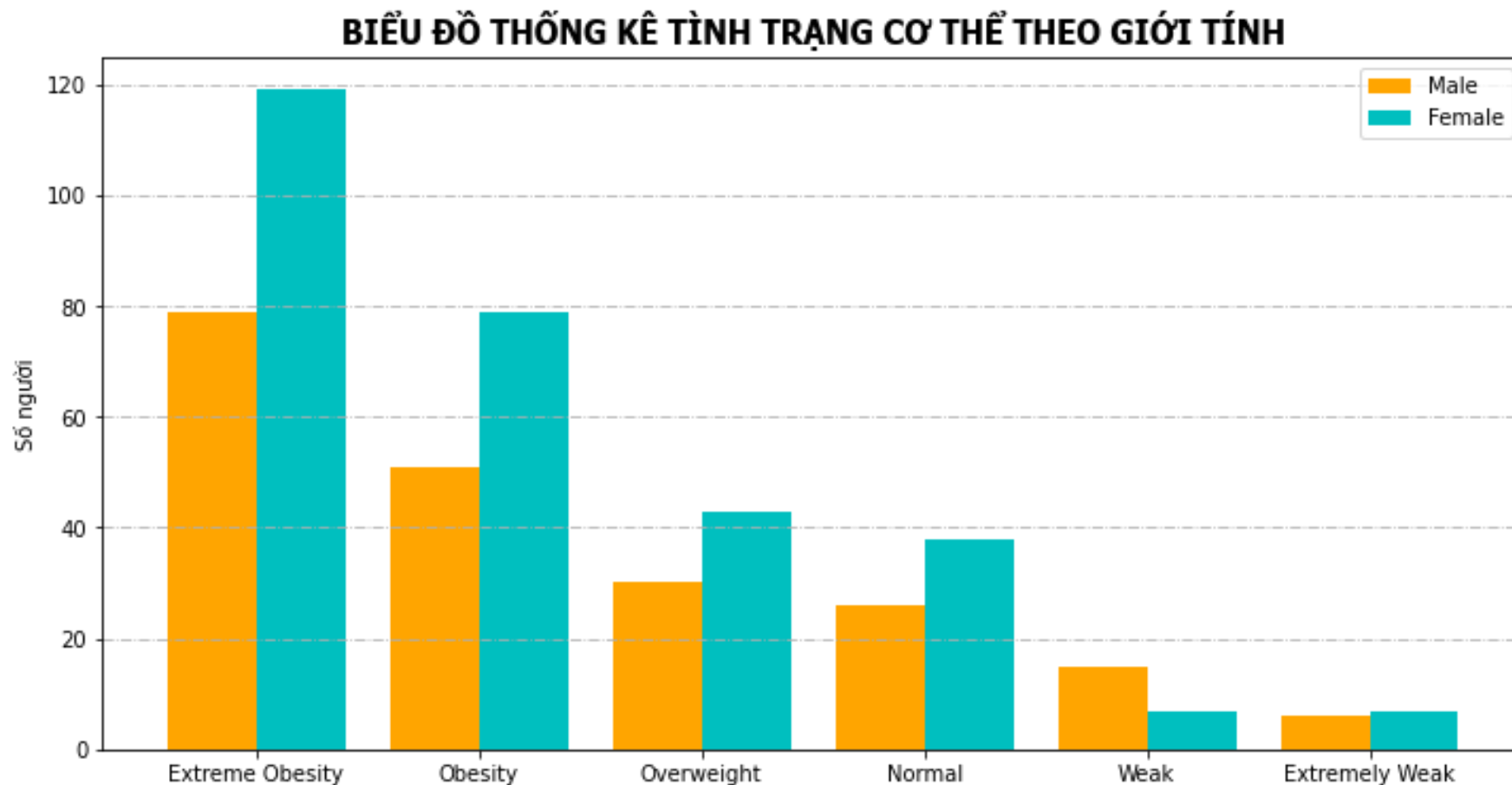
## Yêu cầu 3:

Trực quan hóa phân bố độ tuổi trong tập dữ liệu → cho nhận xét?



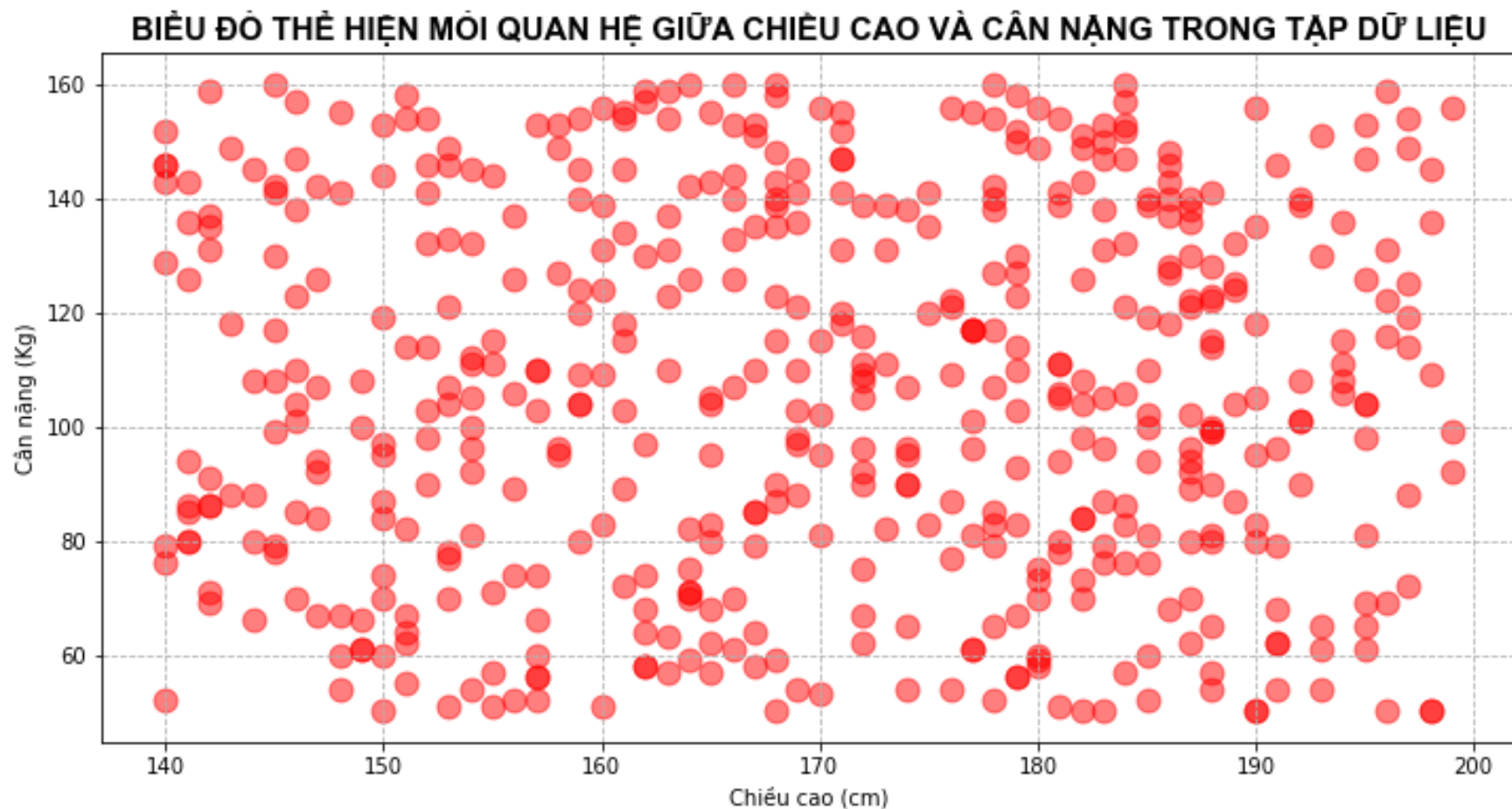
## Yêu cầu 4:

Thống kê chỉ số cơ thể theo từng giới tính và trực quan hóa → cho nhận xét?



## Yêu cầu 5:

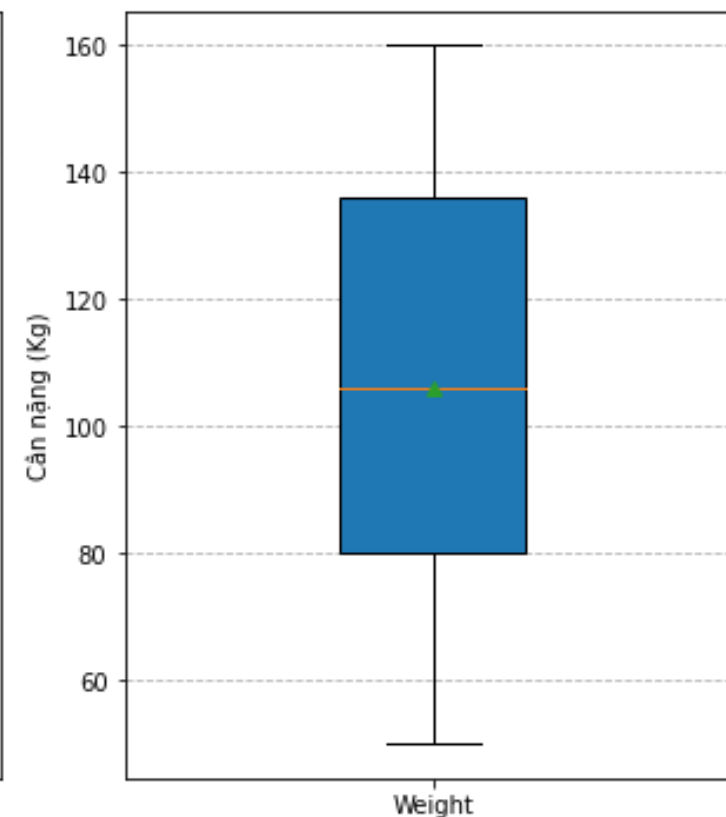
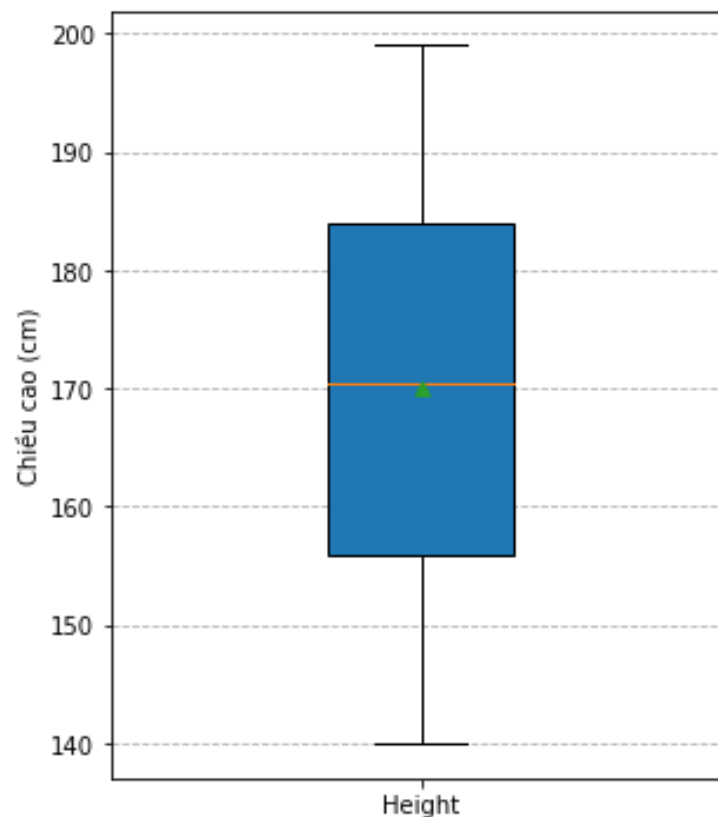
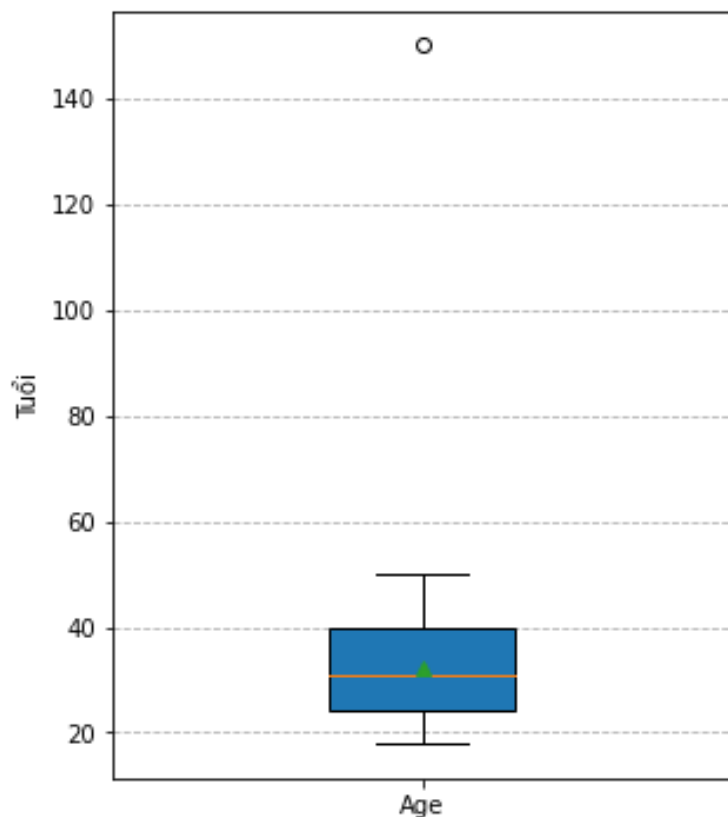
Vẽ biểu đồ thể hiện mối tương quan giữa Height và Weight của 500 người → cho nhận xét?





## Yêu cầu 6:

Sử dụng Boxplot Trực quan hóa dữ liệu thuộc tính Age, Height, Weight → cho nhận xét?



# Thank you!