

Chapter 6 - Exercise 7: Area plot, Box plot

Thực hành vẽ Area plot, Box plot trên 2 tập dữ liệu khác nhau.

Part 1: Area Plot

```
In [1]: 1 import numpy as np
        2 import pandas as pd
        3 import matplotlib.pyplot as plt
```

```
In [2]: 1 # Cho Dữ Liệu Số giờ nắng các tháng trong năm 2016, 2017 tại trạm quan trắc Vũng Tàu
        2 df = pd.DataFrame(
        3     {
        4         'Month': [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12],
        5         'Hours_2017': [183.4, 211.8, 286.4, 287.5, 238.8, 200.3, 187.4,
        6                        233.8, 225.5, 149.1, 180.2, 198.3],
        7         'Hours_2016': [272.8, 254.0, 296.0, 298.0, 240.1, 197.8, 240.3,
        8                        219.5, 212.7, 134.7, 215.3, 109.1]
        9     }
        10 )
```

```
In [3]: 1 # Hiển thị nội dung của df
        2 df
```

Out[3]:

	Month	Hours_2017	Hours_2016
0	1	183.4	272.8
1	2	211.8	254.0
2	3	286.4	296.0
3	4	287.5	298.0
4	5	238.8	240.1
5	6	200.3	197.8
6	7	187.4	240.3
7	8	233.8	219.5
8	9	225.5	212.7
9	10	149.1	134.7
10	11	180.2	215.3
11	12	198.3	109.1

Câu 1: Trên cùng một biểu đồ, hãy vẽ:

- Area plot cho 12 tháng nắng trong năm 2016
- Line plot cho 12 tháng nắng trong năm 2017
- Bạn nhận xét gì về biểu đồ vừa vẽ

Nhấn vào đây để xem kết quả!

Số giờ nắng các tháng trong năm tại trạm quan trắc Vũng Tàu



Part 2: Box plot

Câu 1

```
In [4]: 1 # Cho dữ liệu baseball.csv. Đọc dữ liệu từ baseball.csv và lưu vào biến data,  
2 # hiển thị 10 dòng nội dung đầu của data  
3 data = pd.read_csv('data/baseball.csv', index_col=0)  
4 data.head(10)
```

Out[4]:

	height	weight
0	1.8796	81.646560
1	1.8796	97.522280
2	1.8288	95.254320
3	1.8288	95.254320
4	1.8542	85.275296
5	1.7526	79.832192
6	1.7526	94.800728
7	1.8034	90.718400
8	1.9304	104.779752
9	1.8034	81.646560

```
In [5]: 1 # Cho biết thông tin thống kê chung của data
        2 data.describe()
```

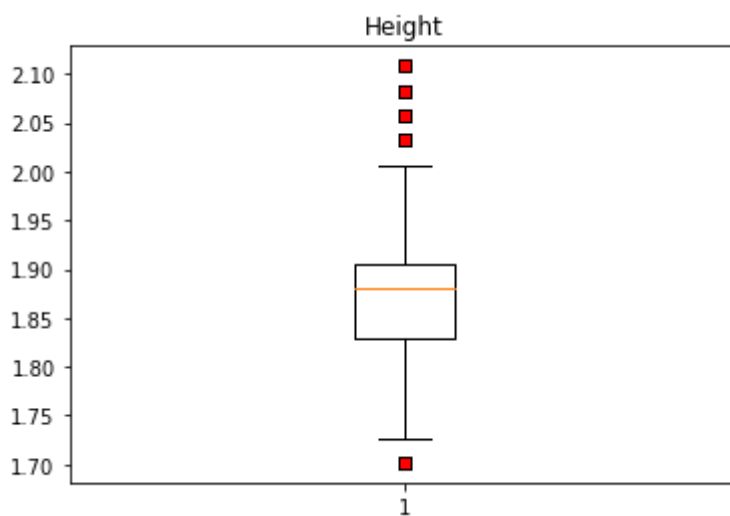
Out[5]:

	height	weight
count	1015.000000	1015.000000
mean	1.871717	91.330191
std	0.058774	9.445198
min	1.701800	68.038800
25%	1.828800	84.368112
50%	1.879600	90.718400
75%	1.905000	97.522280
max	2.108200	131.541680

Câu 2

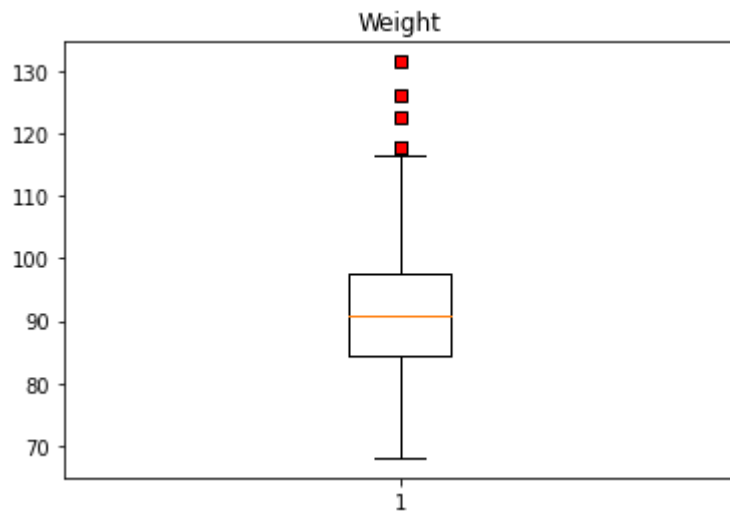
```
In [6]: 1 # Vẽ boxplot cho dữ liệu height
        2 red_square = dict(markerfacecolor = 'r', marker = 's')
        3
```

Nhấn vào đây để xem kết quả!



```
In [7]: 1 # Vẽ boxplot cho dữ liệu weight
        2
```

Nhấn vào đây để xem kết quả!



Câu 3: Kiểm tra xem dữ liệu có outliers hay không? Nếu có thì loại bỏ các outliers. Vẽ lại boxplot

```
In [8]: 1 # Tìm, đếm và in ra các outliers  
        2
```

```
In [9]: 1 # Loại bỏ các outliers  
        2
```

```
In [10]: 1 # Vẽ lại box plot  
         2
```

```
In [ ]: 1
```