


Class period 12

บทที่ 6 การแสดงผลการกระจายของข้อมูล (ต่อ)
Visualize_Data_Distribution


1



Quiz2

- โปรดดูข้อมูลรายได้เฉลี่ยเดือนต่อครัวเรือน จำนวนตามภาค และจังหวัดจาก
- https://data.go.th/dataset/item_7049410f-5bb8-4c75-9e94-112ca18b63e2
- ให้นักศึกษา clean ข้อมูลคอลัมน์ 'ภาคและจังหวัด' และคอลัมน์ที่เป็น missing ออก

2



การแสดงผลการกระจายของข้อมูล Scatter Plot


```

import pandas as pd
from matplotlib import pyplot as plt
df = pd.read_csv('https://raw.githubusercontent.com/pandas-dev/pandas/master/pandas/tests/io/data/csv/iris.csv')
df.head()


สร้าง google colab ใหม่ Visualize_Data_Distribution_02
ดาวน์โหลดข้อมูลดอกไม้ iris

```

3




ดอกไม้ iris




iris setosa

petal sepal



iris versicolor


petal sepal



iris virginica

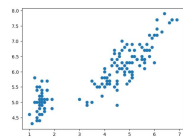
petal sepal

4




Scatter 2D

- คือการมองข้อมูลในตารางแต่ละค่าเป็นจุด 1 จุด และนำจุดมาสร้างกราฟ เพื่อดูการกระจายของข้อมูลระหว่าง 2 nomination (คอลัมน์)
- โดยใช้คำสั่ง plt.scatter(แกนX,แกนY) เช่น
- `plt.scatter(df['PetalLength'],df['SepalLength'])`



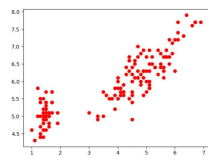
	SepalLength	SepalWidth	PetalLength	PetalWidth	Name
0	5.1	3.5	1.4	0.2	iris setosa
1	4.9	3.0	1.4	0.2	iris setosa
2	4.7	3.2	1.3	0.2	iris setosa
3	4.6	3.1	1.5	0.2	iris setosa
4	5.0	3.6	1.4	0.2	iris setosa

5



Parameter: c กำหนดค่าสีให้ทุกจุด

- สามารถกำหนดค่าสีที่ต้องการให้ทุกจุดได้ แต่ทุกจุดจะสีเดียวกัน เช่น
- `plt.scatter(df['PetalLength'],df['SepalLength'],c='r')`



6

Parameter: c กำหนดค่าสีตามกลุ่มข้อมูลที่ต้องการ

- สามารถกำหนดค่าสีแต่ละจุดตามชื่อดอกไม้แต่ละกลุ่มได้ โดยการแทนที่ข้อมูลชื่อดอกไม้แต่ละกลุ่มด้วยตัวเลข หรือด้วยชื่อสี เช่น
- ใช้คำสั่ง `set(df['Name'])` เพื่อดูชื่อดอกไม้ในคอลัมน์ Name ผลลัพธ์จะได้


```
('Iris-setosa', 'Iris-versicolor', 'Iris-virginica')
```
- ใช้คำสั่ง `replace` เพื่อแทนที่ค่าที่ต้องการด้วยตัวเลขหรือด้วยสี และเก็บตารางที่แทนที่ข้อมูลไว้ในตัวแปรใหม่ เช่น

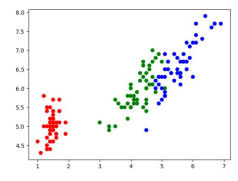

```
df2 = df.replace({'Iris-setosa': 0, 'Iris-versicolor': 1, 'Iris-virginica': 2})
```
- หรือ


```
df2 = df.replace({'Iris-setosa': 'r', 'Iris-versicolor': 'g', 'Iris-virginica': 'b'})
```

7

Parameter: c กำหนดค่าสีตามกลุ่มข้อมูลที่ต้องการ

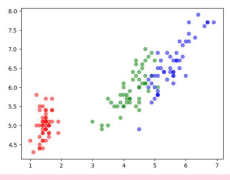
- ให้กำหนด Parameter: c ด้วยค่าในคอลัมน์ Name ของตัวแปรตารางที่แทนที่ค่าแล้ว
- `plt.scatter(df2['PetalLength'], df2['SepalLength'], c=df2['Name'])`



8

alpha กำหนดความโปร่งแสง

- คือค่าความโปร่งแสงของสี มีค่า 0 ถึง 1 โดย 1 คือไม่โปร่งแสง เพื่อให้องเห็นจุดที่ทับกัน (Bubble Chart)
- `plt.scatter(df2['PetalLength'], df2['SepalLength'], c=df2['Name'], alpha=0.5)`



9