



การประมวลผลสารสนเทศ

ด้วยภาษาไพทอน

---

(Information processing in Python)



# จุดประสงค์การเรียนรู้

- ✓ นักเรียนสามารถเขียนโปรแกรมประมวลผลสารสนเทศได้





# เนื้อหาประกอบด้วย



1. การติดตั้งโมดูล pandas
2. โครงสร้างข้อมูล pandas
  - ข้อมูลที่เก็บเป็น Series
  - ข้อมูลที่เก็บเป็น DataFrame
3. ไฟล์ข้อมูลในรูปแบบซีเอสวี
4. การเชื่อมโยงข้อมูล
5. การทำข้อมูลให้เป็นภาพ



## การประมวลผลสารสนเทศ

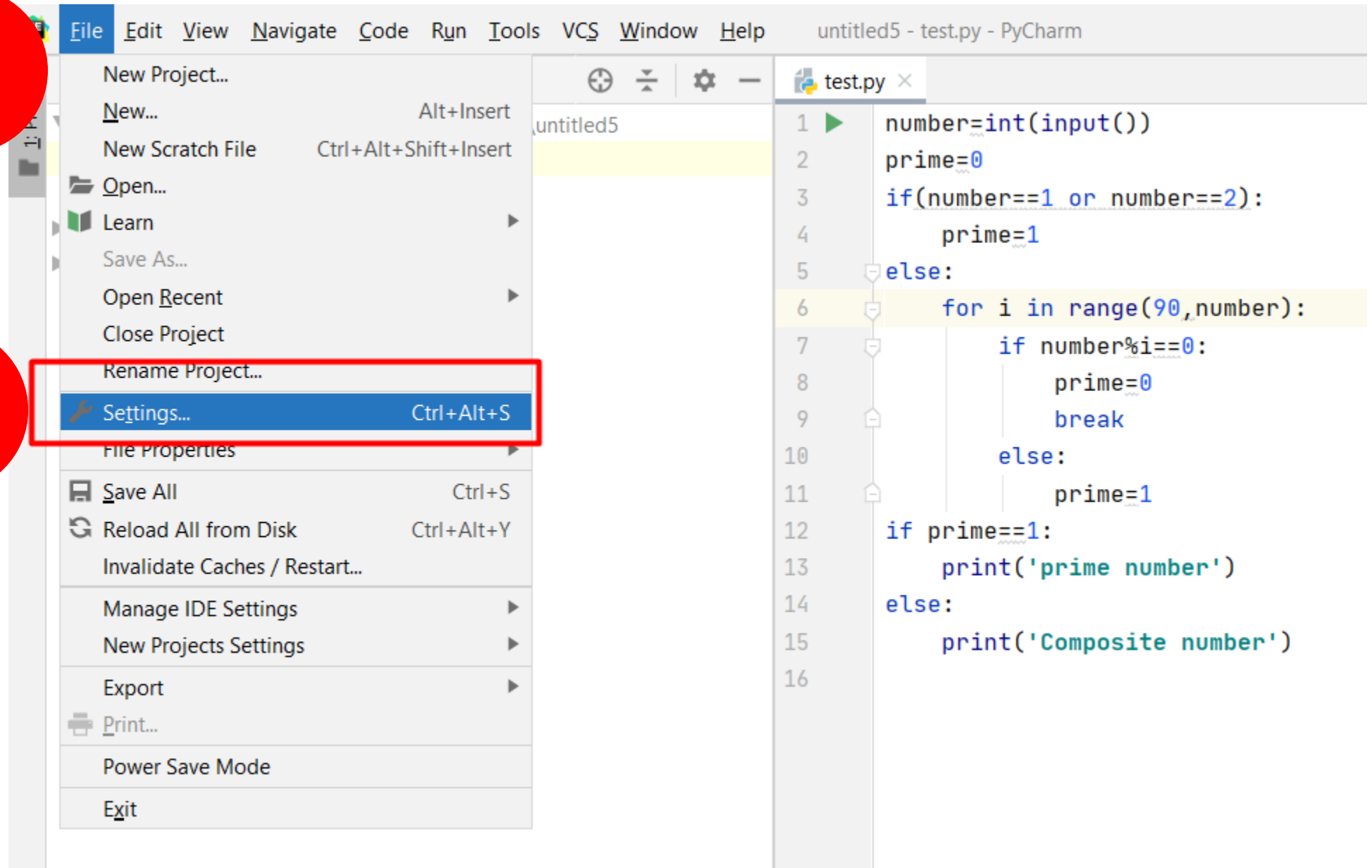
```
mirror_mod = modifier_ob.  
#set mirror object to mirror  
mirror_mod.mirror_object =  
#operation == "MIRROR_X":  
mirror_mod.use_x = True  
mirror_mod.use_y = False  
mirror_mod.use_z = False  
#operation == "MIRROR_Y":  
mirror_mod.use_x = False  
mirror_mod.use_y = True  
mirror_mod.use_z = False  
#operation == "MIRROR_Z":  
mirror_mod.use_x = False  
mirror_mod.use_y = False  
mirror_mod.use_z = True  
  
#selection at the end -add  
mirror_ob.select= 1  
modifier_ob.select=1  
context.scene.objects.active  
("Selected" + str(modifier_ob.name))  
mirror_ob.select = 0  
= bpy.context.selected_objects  
data.objects[one.name].select  
  
print("please select exactly  
  
CLASSES
```

ไพทอนมีโมดูลจำนวนมาก เพื่อใช้จัดการข้อมูล  
ในบทเรียนนี้ขอแนะนำโมดูล **pandas**  
ที่ช่วยทำงานด้านการประมวลผล โดยผู้ใช้  
ต้องติดตั้งโมดูล **Pandas** และ **Numpy**

# 1.การติดตั้งโมดูล Pandas ในโปรแกรม PyCharm

1

2



The screenshot shows the PyCharm IDE interface. The 'File' menu is open, and the 'Settings...' option is highlighted with a red box. A red circle with the number '1' is positioned over the 'File' menu, and another red circle with the number '2' is positioned over the 'Settings...' option. The background shows a Python script named 'test.py' with the following code:

```
1 number=int(input())
2 prime=0
3 if(number==1 or number==2):
4     prime=1
5 else:
6     for i in range(90,number):
7         if number%i==0:
8             prime=0
9             break
10        else:
11            prime=1
12    if prime==1:
13        print('prime number')
14    else:
15        print('Composite number')
16
```



Appearance & Behavior

Keymap

Editor

Plugins

Version Control

Project: untitled5

Project Structure

Python Interpreter

Build, Execution, Deployment

Schemas and DTDs

Tools



Project: untitled5 > Python Interpreter

For current project

Project Interpreter: Python 3.8 (untitled5) C:\Users\Lenovo\PycharmProjects\untitled5\venv\Scripts\python.exe



Package	Version	Latest version	
pip	20.1	20.1	+
setuptools	46.1.3	▲ 46.3.0	-
			▲
			👁



5

Available Packages

Search: pandas

- pandas
- pandas-alive
- pandas-anaphora
- pandas-bamboo
- pandas-bj
- pandas-bokeh
- pandas-cache
- pandas-cacher
- pandas-charm
- pandas-compat
- pandas-connect
- pandas-cub
- pandas-cub-TedPetrou
- pandas-datadict
- pandas-datapackage-reader
- pandas-datareader
- pandas-datareader-gdax
- pandas-ddescribe
- pandas-dedupe
- pandas-diligent
- pandas-ex
- pandas-explode
- pandas-express
- pandas-ext
- pandas-extensions
- pandas-extras
- pandas-f
- pandas-files

Description

Powerful data structures for data analysis, time series, and statistics

Version

1.0.3

<https://pandas.pydata.org>

Specify version 1.0.3

Options

6

Install Package    Manage Repositories

# วิดีโอเกี่ยวกับการติดตั้งโมดูล `pandas` ในโปรแกรม `pycharm`

ก็อปปีลิงค์ไปวางในเว็บเบราว์เซอร์

<https://youtu.be/aTspc95ITVU>





## 2. โครงสร้างข้อมูลของโมดูล **pandas**

ข้อมูลที่เก็บใน **pandas** จะอยู่ในรูปแบบชุดข้อมูลที่เป็น

**Series**  
(ซีรีส์)

**DataFrame**  
(ดาต้าเฟรม)

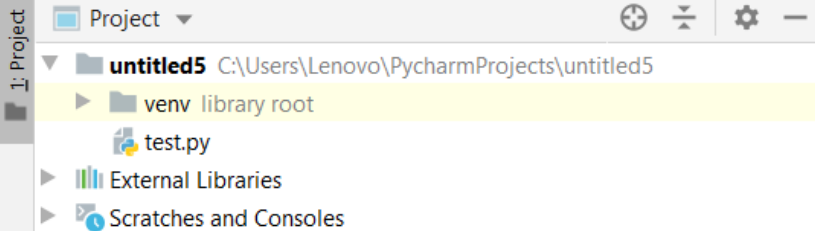


## 2.1) ข้อมูลที่เก็บเป็น Series

ข้อมูลที่เก็บเป็น **series** จะมีลักษณะคล้ายลิสต์ของข้อมูลหนึ่งมิติ

ตัวอย่าง การใช้ **series** ในการเก็บคะแนนสอบวิชาวิทยาการคำนวณ  
ของนักเรียนชั้น ม.3/1 จำนวน 20 คน

```
1 import pandas as pd
2
3 s=pd.Series([18,20,15,6,9,10,20,19,17,13,16,7,14,15,20,4,12,17,19,18])
4 print(s)
```



```
test.py x
1 import pandas as pd
2
3 s=pd.Series([18,20,15,6,9,10,20,19,17,13,16,7,14,15,20,4,12,17,19,18])
4 print(s)
5
```

Run: test x

0	18
1	20
2	15
3	6
4	9
5	10
6	20
7	19
8	17
9	13
10	16
11	7
12	14
13	15
14	20
15	4
16	12
17	17
18	19
19	18

dtype: int64

ผลลัพธ์ที่ได้

ชื่อไฟล์ series.py

Process finished with exit code 0



## 2.2) ข้อมูลที่เก็บเป็น DataFrame

ข้อมูลที่เก็บเป็น DataFrame จะมีลักษณะคล้ายตาราง (2 มิติขึ้นไป)

### ตัวอย่าง การใช้งาน DataFrame

```
import pandas as pd
```

```
df=pd.DataFrame([[1,'a'], [2,'b'], [3,'c']])  
print(df)
```

The screenshot shows the PyCharm IDE interface. The top menu bar includes File, Edit, View, Navigate, Code, Run, Tools, VCS, Window, and Help. The title bar indicates the current file is 'untitled5 - D:\งานการสอน\ปีการศึกษา 2563\Powerpoint ม.3\เทอม 1\ไฟล์ python\2.Da'. The Project tool window on the left shows a project named 'untitled5' with a virtual environment 'venv library root' and a file 'test.py'. The main editor window displays the code in '2.DataFrame.py':

```
1 import pandas as pd
2 df=pd.DataFrame([[1, 'a'], [2, 'b'], [3, 'c']])
3 print(df)
4
```

The Run tool window at the bottom shows the execution output for '2.DataFrame.py':

```
0 1
0 1 a
1 2 b
2 3 c

Process finished with exit code 0
```

ผลลัพธ์ที่ได้

ชื่อไฟล์ DataFrame.py



### 3.ไฟล์ข้อมูลในรูปแบบซีเอสวี

ไฟล์รูปแบบซีเอสวี หรือ CSV (comma separated values) คือไฟล์ข้อมูลที่แต่ละตัวในแถวเดียวกัน ถูกคั่นด้วยเครื่องหมายจุลภาค (,) โดยอาจมีแถวแรกสุดเป็นชื่อของคอลัมน์ หรือหัวตาราง

Website Name	Visitor/Day
imotab.com	5,500
thailandseo.com	1,200
compact-dslr.com	2,500
technointrend.com	1,500

```
1 Website Name,Visitor/Day
2 imotab.com," 5,500 "
3 thailandseo.com," 1,200 "
4 compact-dslr.com," 2,500 "
5 technointrend.com," 1,500 "
6
```

CSV File

ตัวอย่าง ข้อมูลของนักเรียนชั้น ม.3/1 จำนวน 20 คน ประกอบด้วย  
เลขประจำตัวนักเรียน เพศ (1 แทนเพศชาย และ 2 แทนเพศหญิง)  
คะแนนสอบวิชาวิทยาการคำนวณ และส่วนสูง

เลขประจำตัว	เพศ	Cs	height
5884	2	18	158
5885	2	20	164
5887	1	15	150
5889	1	6	168
5890	2	9	153
5893	1	10	172
5895	1	20	162
5896	1	19	157
5897	2	17	169
5900	2	13	174
5903	1	16	150
5905	2	7	150
5906	1	14	169
5908	1	15	175
5910	1	20	173
5912	1	4	173
5915	1	12	175
5916	2	17	154
5919	1	19	174
5921	2	18	158

The screenshot shows the PyCharm IDE interface. The top menu bar includes File, Edit, View, Navigate, Code, Run, Tools, VCS, Window, and Help. The title bar indicates the file path: untitled5 - D:\งานการสอน\ปีการศึกษา 2563\Powerpoint ม.3\เทอม 1\ไฟล์ python\DataCSV.py - PyChar. The Project tool window on the left shows the project structure for 'untitled5', including a 'venv' directory and 'test.py'. The Editor tool window displays the code in 'DataCSV.py':

```
1 import pandas as pd
2 stdData=pd.read_csv('studentData.csv',delimiter=',')
3 print(stdData)
4
```

The Run tool window shows the output of the script, which is a CSV file with the following data:

	เลขประจำตัว	เพศ	Cs	height
0	5884	2	18	158
1	5885	2	20	164
2	5887	1	15	150
3	5889	1	6	168
4	5890	2	9	153
5	5893	1	10	172
6	5895	1	20	162
7	5896	1	19	157
8	5897	2	17	169
9	5900	2	13	174
10	5903	1	16	150
11	5905	2	7	150
12	5906	1	14	169
13	5908	1	15	175
14	5910	1	20	173
15	5912	1	4	173
16	5915	1	12	175
17	5916	2	17	154
18	5919	1	19	174
19	5921	2	18	158

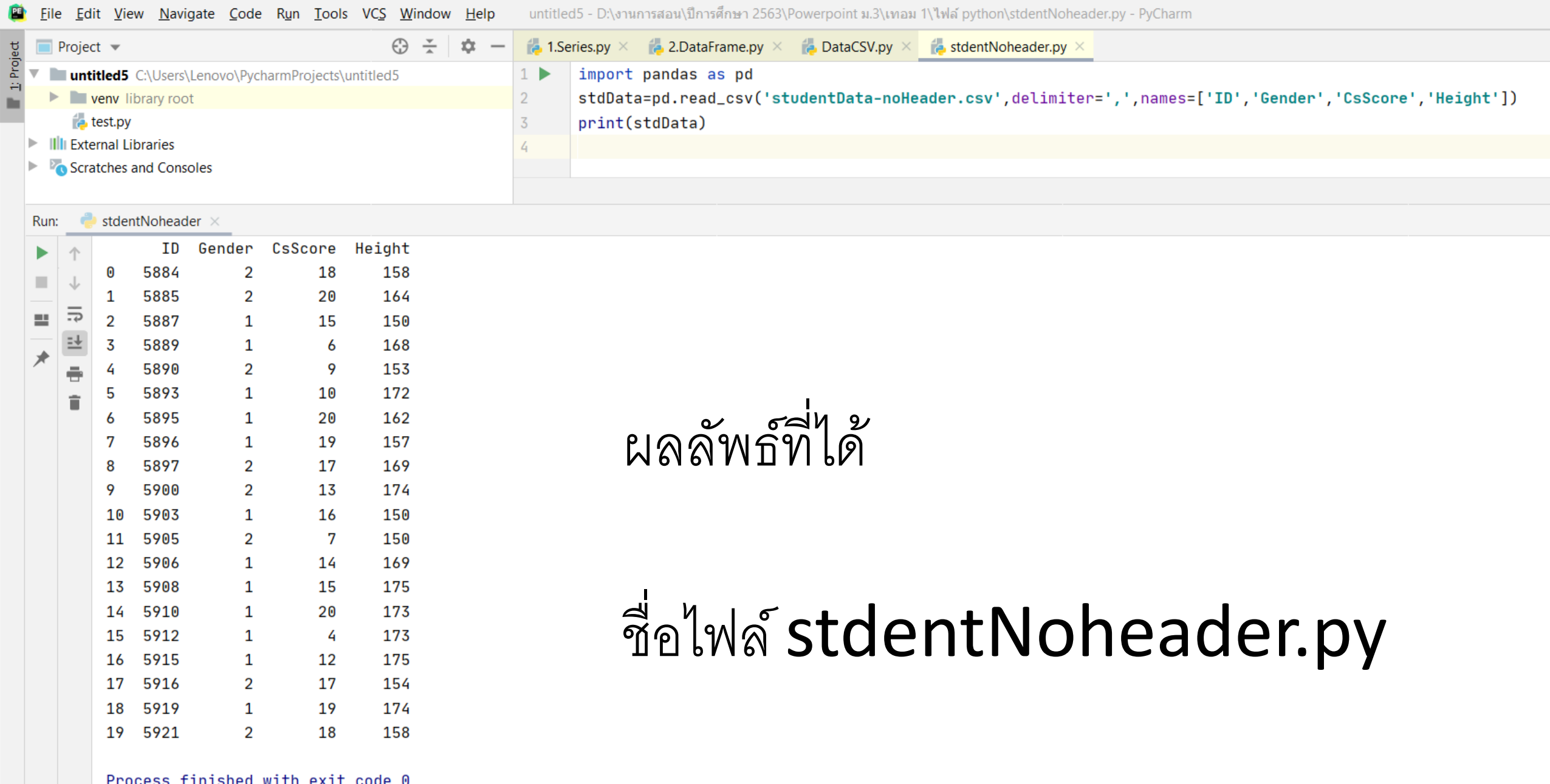
At the bottom of the Run tool window, it states: "Process finished with exit code 0".

ผลลัพธ์ที่ได้

ชื่อไฟล์ DataCSV.py



# กรณีที่ไฟล์ CSV ไม่มีแถวที่เป็นชื่อคอลัมน์หรือหัวตาราง ผู้ใช้สามารถระบุหัวตารางในการแสดงผลได้



The screenshot shows the PyCharm IDE interface. The top toolbar includes File, Edit, View, Navigate, Code, Run, Tools, VCS, Window, and Help. The main editor window displays the following Python code in `stdntNoheader.py`:

```
1 import pandas as pd
2 stdData=pd.read_csv('studentData-noHeader.csv',delimiter=',',names=['ID','Gender','CsScore','Height'])
3 print(stdData)
4
```

The Run console shows the output of the script, which is a DataFrame with 20 rows and 5 columns: ID, Gender, CsScore, and Height. The data is as follows:

	ID	Gender	CsScore	Height
0	5884	2	18	158
1	5885	2	20	164
2	5887	1	15	150
3	5889	1	6	168
4	5890	2	9	153
5	5893	1	10	172
6	5895	1	20	162
7	5896	1	19	157
8	5897	2	17	169
9	5900	2	13	174
10	5903	1	16	150
11	5905	2	7	150
12	5906	1	14	169
13	5908	1	15	175
14	5910	1	20	173
15	5912	1	4	173
16	5915	1	12	175
17	5916	2	17	154
18	5919	1	19	174
19	5921	2	18	158

At the bottom of the Run console, it says "Process finished with exit code 0".

ผลลัพธ์ที่ได้

ชื่อไฟล์ `stdntNoheader.py`

```
1 ▶ import pandas as pd
2 stdData=pd.read_csv('studentData-noHeader.csv',delimiter=',',names=['ID','Gender','CsScore','Height'])
3 print(stdData)
4
```

การแสดงผลข้อมูลใน DataFrame ออกทางจอภาพ เราสามารถใช้ฟังก์ชัน `print()` แสดงผลลัพธ์ในรูปแบบต่างๆได้

```
print(stdData)
```

```
print(stdData.head(2))
```

```
print(stdData.tail(2))
```

```
print(stdData[['ID', 'CsScore']])
```

```
print(stdData[['ID', 'CsScore']].head(2))
```

```
print(stdData['CsScore'])
```

```
print(stdData[stdData.Gender==2])
```

```
print(stdData[stdData.CsScore>=16])
```

## การนับจำนวนข้อมูล

ใช้ฟังก์ชัน `count( )`

```
print(stdData.count())
```

## การหาค่าเฉลี่ย

ใช้ฟังก์ชัน `mean( )`

```
print(stdData.CsScore.mean())
```

## การหาค่าเฉลี่ยแยกกลุ่ม

ใช้ฟังก์ชัน `groupby( )`

```
print(stdData.groupby('Gender').mean())
```

## การจัดเรียงข้อมูล

ใช้ฟังก์ชัน

`sort_values( )`

```
print(stdData.sort_values(by=['Height']))
```



## 4. การเชื่อมโยงข้อมูล

การเชื่อมโยงข้อมูล 2 ไฟล์

sciScore.csv		
	A	B
1	เลขประจำตัว	Science1
2	5897	8
3	5919	16
4	5910	15
5	5896	7
6	5900	18
7	5893	4
8	5908	10
9	5887	9
10	5885	17
11	5921	4
12	5912	3
13	5916	4
14	5915	11
15	5905	12
16	5889	8
17	5895	16
18	5890	14
19	5906	4
20	5903	7
21	5884	14

studentData.csv				
	A	B	C	D
1	เลขประจำตัว	เพศ	Cs	height
2	5884	2	18	158
3	5885	2	20	164
4	5887	1	15	150
5	5889	1	6	168
6	5890	2	9	153
7	5893	1	10	172
8	5895	1	20	162
9	5896	1	19	157
10	5897	2	17	169
11	5900	2	13	174
12	5903	1	16	150
13	5905	2	7	150
14	5906	1	14	169
15	5908	1	15	175
16	5910	1	20	173
17	5912	1	4	173
18	5915	1	12	175
19	5916	2	17	154
20	5919	1	19	174
21	5921	2	18	158

```
File Edit View Navigate Code Run Tools VCS Window Help
untitled5 - D:\งานการสอน\ปีการศึกษา 2563\Powerpoint ม.3\เทอม 1\ไฟล์ python\connet2file.py - PyCharm

Project
└─ untitled5 C:\Users\Lenovo\PycharmProjects\untitled5
   └─ venv library root
      └─ test.py
   └─ External Libraries
   └─ Scratches and Consoles

1 Series.py x 2.DataFrame.py x DataCSV.py x stdentNoheader.py x connet2file.py x
1 import pandas as pd
2
3 stdData=pd.read_csv('studentData.csv',delimiter=',')
4 sciData=pd.read_csv('sciScore.csv',delimiter=',')
5
6 newData=stdData.merge(sciData)
7 print(newData)
8
```

Run: connet2file x

0	5884	2	18	158	14
1	5885	2	20	164	17
2	5887	1	15	150	9
3	5889	1	6	168	8
4	5890	2	9	153	14
5	5893	1	10	172	4
6	5895	1	20	162	16
7	5896	1	19	157	7
8	5897	2	17	169	8
9	5900	2	13	174	18
10	5903	1	16	150	7
11	5905	2	7	150	12
12	5906	1	14	169	4
13	5908	1	15	175	10
14	5910	1	20	173	15
15	5912	1	4	173	3
16	5915	1	12	175	11
17	5916	2	17	154	4
18	5919	1	19	174	16
19	5921	2	18	158	4

Process finished with exit code 0

ผลลัพธ์ที่ได้

ชื่อไฟล์ connect2file.py

Project

- untitled5 C:\Users\Lenovo\PycharmProjects\untitled5
  - venv library root
  - test.py
  - External Libraries
  - Scratches and Consoles

```

1 import pandas as pd
2
3 stdData=pd.read_csv('studentData-noHeader.csv',delimiter=',',names=['ID','Gender','CsScore','Height'])
4 print(stdData)
5 sciData=pd.read_csv('sciScore.csv',delimiter=',')
6
7 newData=stdData.merge(sciData,left_on='ID',right_on='เลขประจำตัว')
8 print(newData)
9

```

การเชื่อมโยงข้อมูลกรณีชื่อคอลัมน์หลักไม่ตรงกัน

Run: connect2file-1

	ID	Gender	CsScore	Height	เลขประจำตัว	Science1
0	5884	2	18	158	5884	14
1	5885	2	20	164	5885	17
2	5887	1	15	150	5887	9
3	5889	1	6	168	5889	8
4	5890	2	9	153	5890	14
5	5893	1	10	172	5893	4
6	5895	1	20	162	5895	16
7	5896	1	19	157	5896	7
8	5897	2	17	169	5897	8
9	5900	2	13	174	5900	18
10	5903	1	16	150	5903	7
11	5905	2	7	150	5905	12
12	5906	1	14	169	5906	4
13	5908	1	15	175	5908	10
14	5910	1	20	173	5910	15
15	5912	1	4	173	5912	3
16	5915	1	12	175	5915	11
17	5916	2	17	154	5916	4
18	5919	1	19	174	5919	16
19	5921	2	18	158	5921	4

ผลลัพธ์ที่ได้

ชื่อไฟล์ connect2file-1.py



## 5.การทำข้อมูลให้เป็นภาพ

ต้องติดตั้งโมดูล **matplotlib** ก่อน  
วิธีติดตั้งเหมือนกันกับการติดตั้งโมดูล **pandas**

การนำเสนอผลคะแนนด้วยฮิสโทแกรม

การนำเสนอผลคะแนนด้วยแผนภาพการกระจาย



Project

- untitled5 C:\Users\Lenovo\PycharmProjects\untitled5
  - venv library root
  - test.py
- External Libraries
- Scratches and Consoles

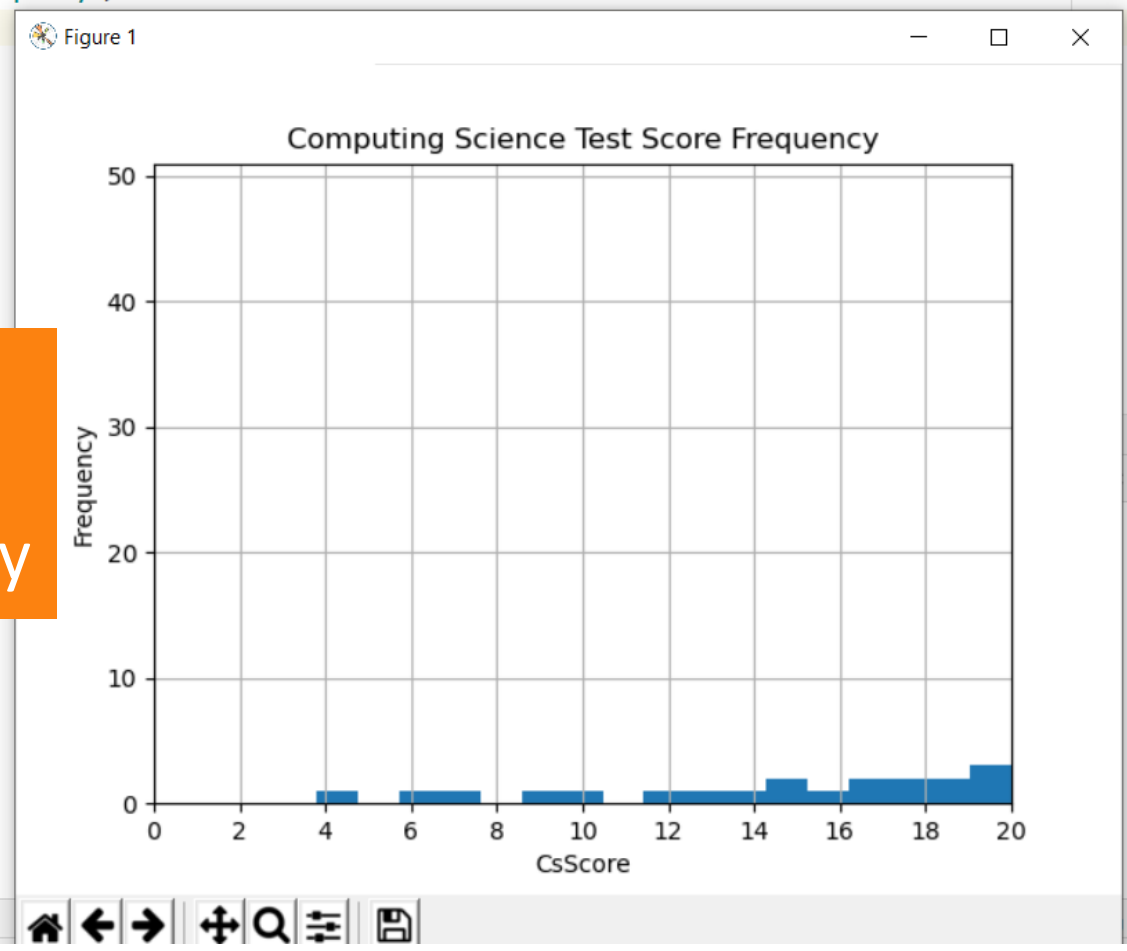
```
1 import pandas as pd
2 import matplotlib.pyplot as plt
3
4 stdData=pd.read_csv('studentData-noHeader.csv',delimiter=',',names=['ID','Gender','CsScore','Height'])
5 sciData=pd.read_csv('sciScore.csv',delimiter=',')
6
7 newData=stdData.merge(sciData,left_on='ID',right_on='เลขประจำตัว',how='inner')
8 print(newData)
9 newData.CsScore.hist(range=[0,20],bins=21)
10 plt.title('Computing Science Test Score Frequency')
11 plt.xlabel('CsScore')
12 plt.ylabel('Frequency')
13 plt.axis([0,20,0,51])
14 plt.xticks(range(0,21,2))
15 plt.show()
16
```

การนำเสนอผลคะแนนด้วยฮิสโทแกรม

ผลลัพธ์  
ไฟล์ histogram.py

Run: histogram

10	5903	1	16	150	5903	7
11	5905	2	7	150	5905	12
12	5906	1	14	169	5906	4
13	5908	1	15	175	5908	10
14	5910	1	20	173	5910	15
15	5912	1	4	173	5912	3
16	5915	1	12	175	5915	11
17	5916	2	17	154	5916	4
18	5919	1	19	174	5919	16
19	5921	2	18	158	5921	4



Project

- untitled5
  - venv library root
  - test.py
  - External Libraries
  - Scratches and Consoles

```

1 import pandas as pd
2 import matplotlib.pyplot as plt
3
4 stdData=pd.read_csv('studentData-noHeader.csv',delimiter=',',names=['ID','Gender','CsScore','Height'])
5 sciData=pd.read_csv('sciScore.csv',delimiter=',')
6
7 newData=stdData.merge(sciData,left_on='ID',right_on='เลขประจำตัว',how='inner')
8 print(newData)
9
10 plt.scatter(newData.CsScore,newData.Science1)
11 plt.xlabel('Computer Science score')
12 plt.ylabel('Frequency')
13
14 for i in range(0,20):
15     plt.text(newData.CsScore[i],newData.Science1[i],newData.ID[i])
16 plt.title('Computing Science score & Science score relationship')
17 plt.show()
18

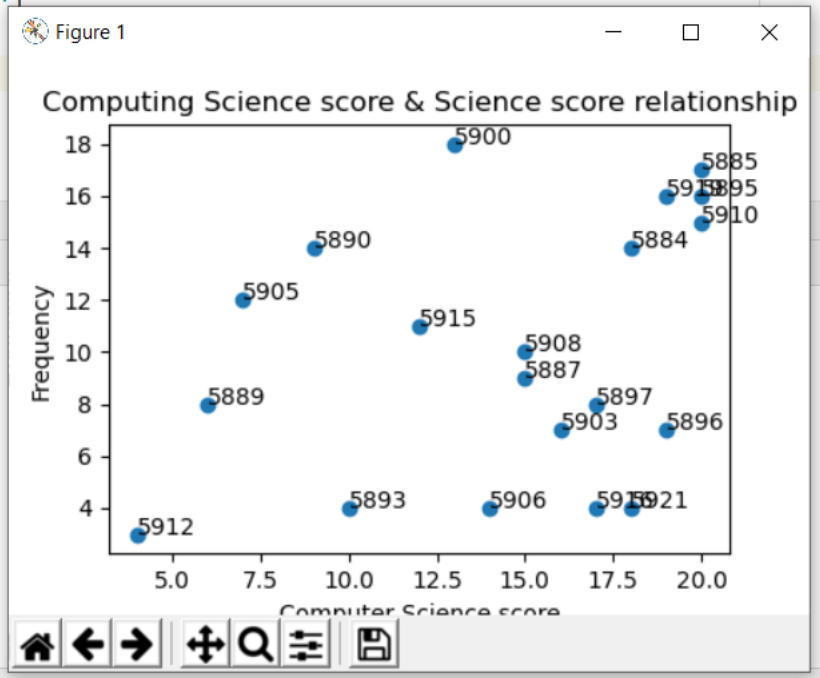
```

การนำเสนอผลคะแนน  
ด้วยแผนภาพการกระจาย

Run: star

10	5903	1	16	150	5903	7
11	5905	2	7	150	5905	12
12	5906	1	14	169	5906	4
13	5908	1	15	175	5908	10
14	5910	1	20	173	5910	15
15	5912	1	4	173	5912	3
16	5915	1	12	175	5915	11
17	5916	2	17	154	5916	4
18	5919	1	19	174	5919	16
19	5921	2	18	158	5921	4

ผลลัพธ์  
ไฟล์ star.py





# แบบฝึกหัดท้ายบท

1.กิจกรรมที่ 2.1