Data Science Homework 1 – Basic Python

- **Objective**
  - Get familiar with python syntax
  - Learn to use python libraries
  - Text data parsing with python
  - Data visualization with python

- **Data**

class	male population	male smoke percentage	female population	female smoke percentage
Education level				
elementary school and below	1596	25.3	2781	1.7
junior high	1264	49.6	1498	10.6
senior high	3136	28.7	3734	6.5
university	2881	11.7	3249	1
graduate school and above	964	4.6	659	0
Average monthly income				
20000 and below	3737	20.2	6382	2.5
20001-40000	3431	33.9	4278	6.8
40001 and above	3615	35.6	2227	5.5
Working environment				
indoor	3739	32.4	4732	3.6
outdoor	1773	40.9	635	9
unemployed	1595	29.5	4114	4.7

  - A survey of smoking percentage of the population
  - The first raw is the title of each column
  - The raw with only first column marks the start of a class of data
  - The data is at https://ceiba.ntu.edu.tw/course/481ea4/hw1_data.csv

- **Tasks**
  - Read the data from the url
  - Parse the data
  - Visualize the data with matplotlib.pyplot

Note that sys, ssl, matplotlib, and urllib are the only libraries that are allowed in HW1.

There are 3 types of chart in this homework, the **line chart**, the **bar chart**, and the **pie chart**. For the line chart and the bar chart, there should be three lines/bars which represent the male smoking percentage, female smoking percentage and total smoking percentage respectively.

You should show in the chart the title of the chart, the label of the axes and the legend of the chart. Also, you should show the percentage value of each data point/bar. An example of the line/bar chart for the Education level is shown in Fig. 1 and 2.

For the pie chart, you should show the proportion of different classes in the smoking population. We assume that the distribute of population in different classes are based on real proportion.

You should show the title of the figure, the class and the percentage for each segment in the pie chart. An example of the pie chart for education level is shown in Fig. 3.

● Format

1. You should write a single hw1.py script. The command line arguments indicate the class of data and the type of chart to plot.
2. For the class of data, **E** represents the education level, **A** the average monthly income, and **W** the working environment.
3. For the type of chart, **l** (lower case L) marks the line chart, **b** the bar chart, and **p** the pie chart.

The command line arguments should be in the format of

–(class of data)(type of chart)

For example, **-Ab** represent the bar chart of the average monthly income data and **-Wp** represent the pie chart of the working environment data.

The arguments can be cascaded, for example,

python hw1.py –Ab –Wp

First shows the bar chart. After the user close the bar chart, it shows the pie chart.

● Submission

You should submit a zip file that contains only hw1.py and a readme which briefly describes your script.

● Grading

1. Plagiarism results in the fail of the course.
2. Correct submission format worth **5pt.**
3. A python script with no error worth **5pt**.
4. Parse the command line arguments correctly worth **15pt**.
5. Parse the data correctly worth **20pt**.
6. Each type of chart worth **15pt**.
7. A readme file that briefly describes your program worth **10pt**.
8. We accept late submission for at most 2 days.
9. The late submission penalty is 15pt per day.

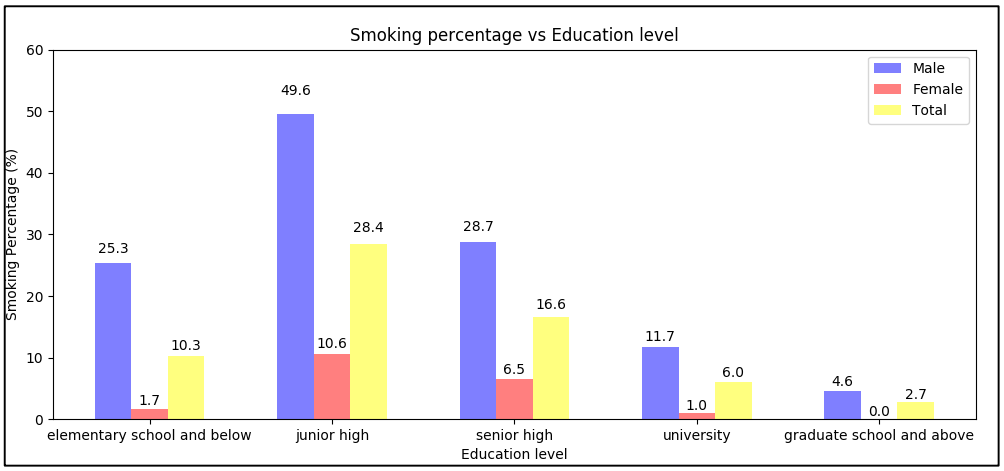

Fig. 1. The bar chart

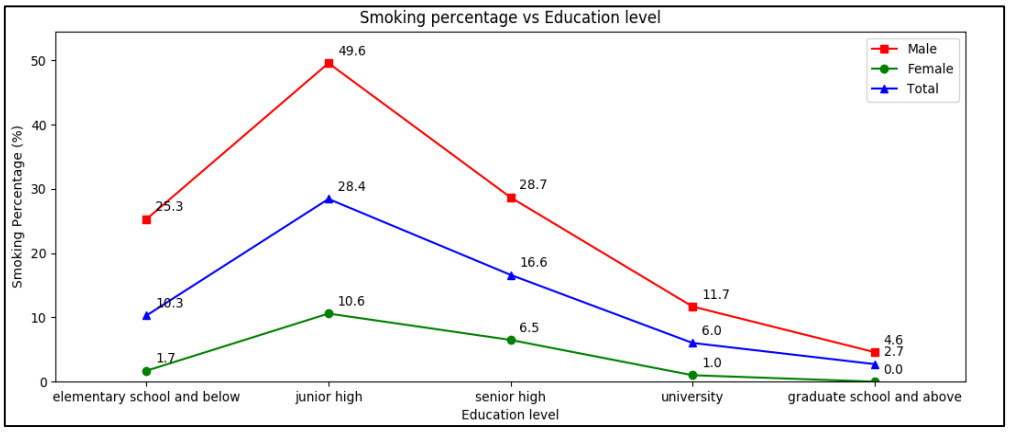

Fig. 2. The line chart

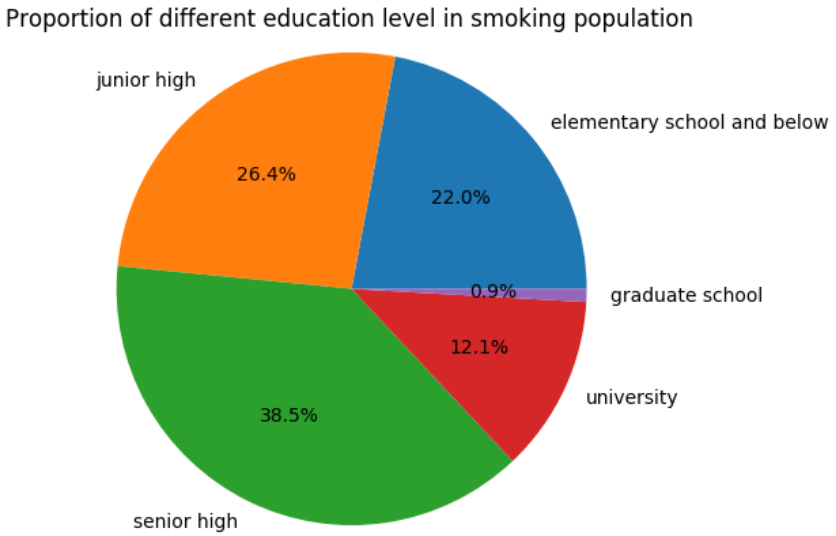

Fig. 3. The pie chart