

# Data Engineering for Python

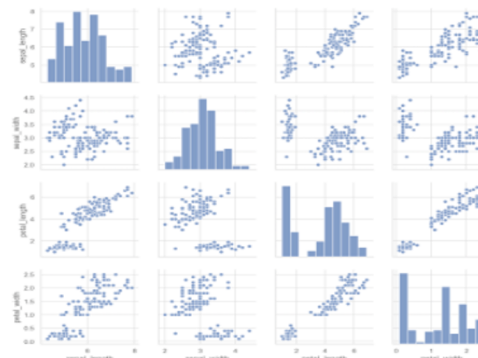
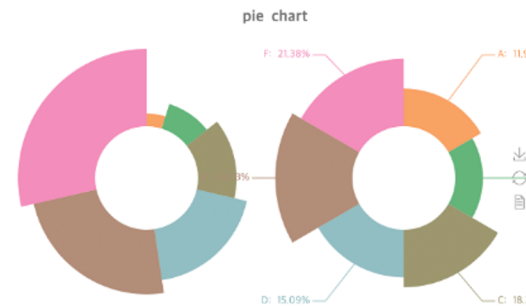
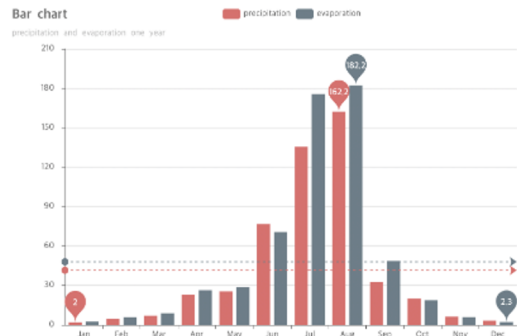
Data Visualization

EDA

Data Preprocessing

- Image Data
- Time Series Data
- Etc ..

Machine Learning algorithms



**Woody K**  
Contact: [woody35545@gmail.com](mailto:woody35545@gmail.com)  
Github: <https://github.com/woody35545>  
CNU, Computer Science and Engineering



# Data Visualization for Python

# Basic Data Structures

Numpy & Pandas



# Numpy



Python의 기본 Data Structure인 list 보다 연산 속도면에서 우수할 뿐 아니라,  
Matrix 연산이나 수치 데이터 처리에 유용한 함수들 다수 제공

List로 Data Frame 생성

```
[2] data_list = [['Steve', '23', 'male'],  
                ['Alex', '25', 'male'],  
                ['Ann', '20', 'female']]
```

```
[3] data_list  
[['Steve', '23', 'male'], ['Alex', '25', 'male'], ['Ann', '20', 'female']]
```

```
[6] dataframe = pd.DataFrame(data_list)
```

```
dataframe
```

	0	1	2
0	Steve	23	male
1	Alex	25	male
2	Ann	20	female

Ref: <https://compmath.korea.ac.kr/appmath/NumpyBasics.html>



Woody K  
Contact: [woody35545@gmail.com](mailto:woody35545@gmail.com)  
Github: <https://github.com/woody35545>  
CNU, Computer Science and Engineering



Python의 기본 Data Structure인 list 보다 연산 속도면에서 우수할 뿐 아니라,  
Matrix 연산이나 수치 데이터 처리에 유용한 함수들 다수 제공

List로 Data Frame 생성

```
[2] data_list = [['Steve', '23', 'male'],  
                ['Alex', '25', 'male'],  
                ['Ann', '20', 'female']]
```

```
[3] data_list  
  
[['Steve', '23', 'male'], ['Alex', '25', 'male'], ['Ann', '20', 'female']]
```

```
[6] dataframe = pd.DataFrame(data_list)
```

```
dataframe
```

	0	1	2
0	Steve	23	male
1	Alex	25	male
2	Ann	20	female

# Data Visualization for Python

# Data Visualization

**The Need of Data Visualization** Page 3

**Well-Known Libraries** Page.21

**How to use** Page.21

**Case Study** Page.21



Woody K  
Contact: [woody35545@gmail.com](mailto:woody35545@gmail.com)  
Github: <https://github.com/woody35545>  
CNU, Computer Science and Engineering

➤ **The Need of Data Visualization** Page 3

➤ **Well-Known Libraries** Page.21

➤ **How to use** Page.21  
Matplotlib, Seaborn

➤ **Case Study** Page.21



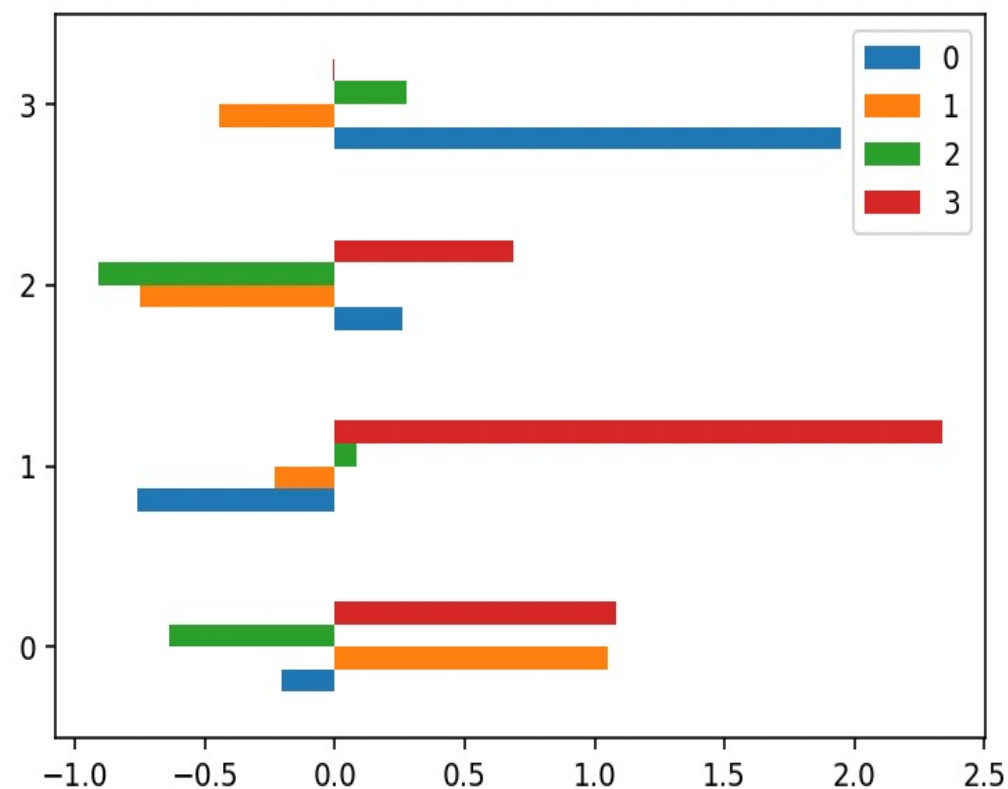
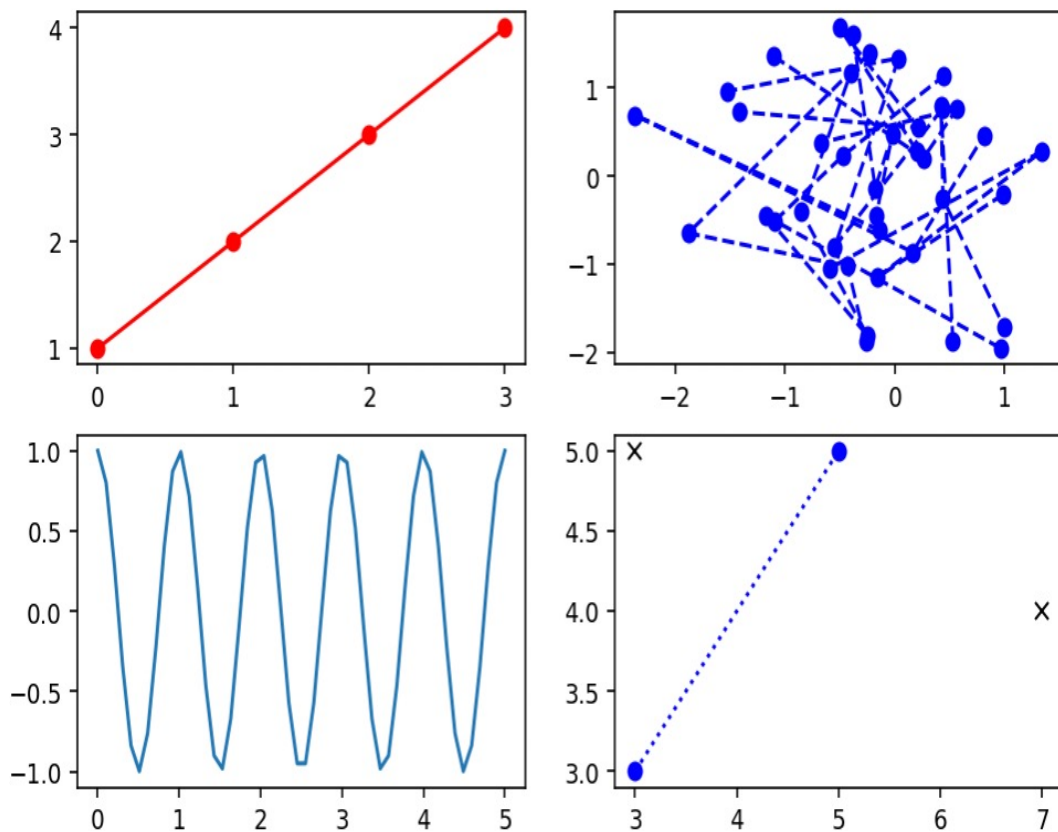


# Data Visualization for Python

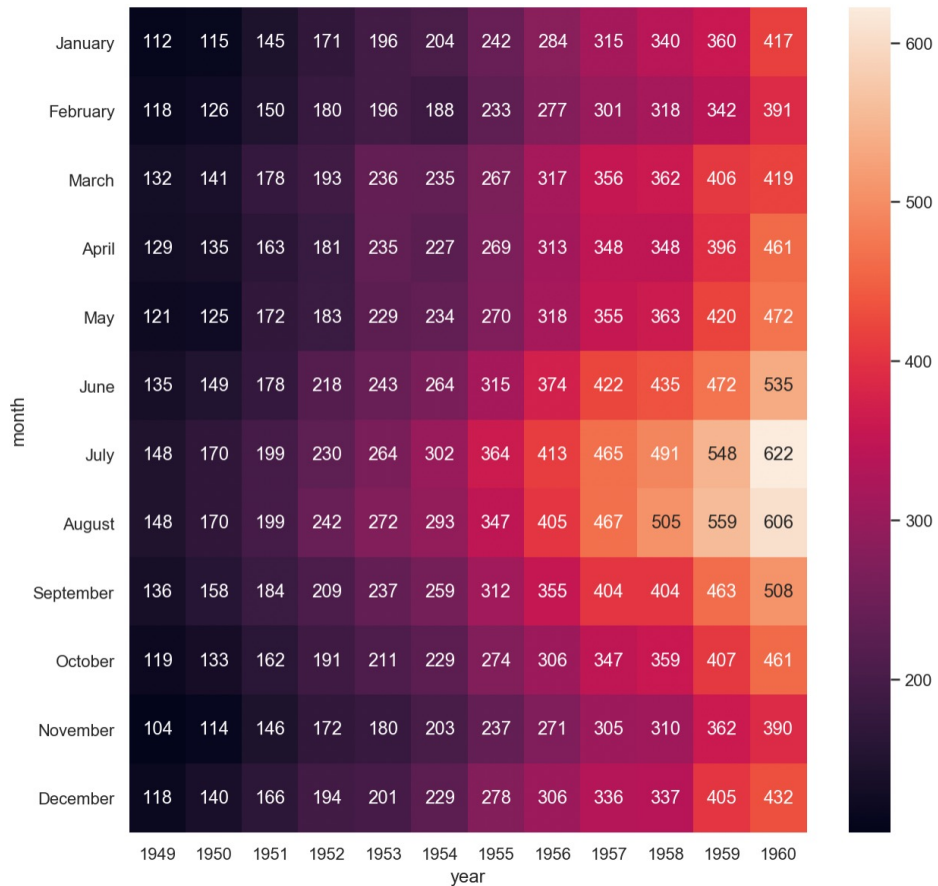
## 1. Well-Known Libs



## Well-Known Libraries



# Well-Known Libraries



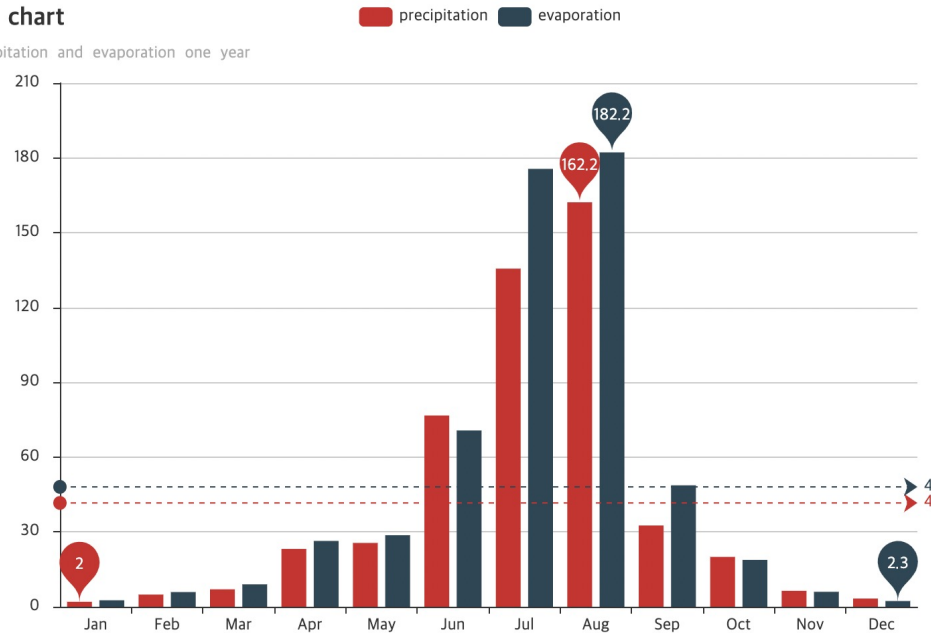
**Woody K**  
Contact: [woody35545@gmail.com](mailto:woody35545@gmail.com)  
Github: <https://github.com/woody35545>  
CNU, Computer Science and Engineering

# Well-Known Libraries

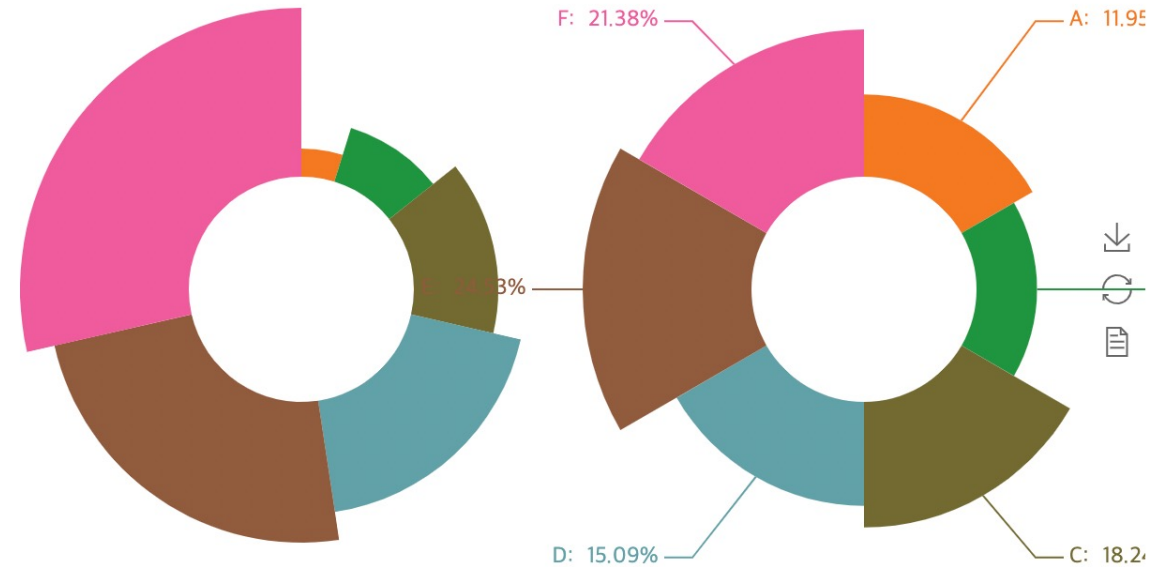
## PyeChart

Bar chart

precipitation and evaporation one year



pie chart

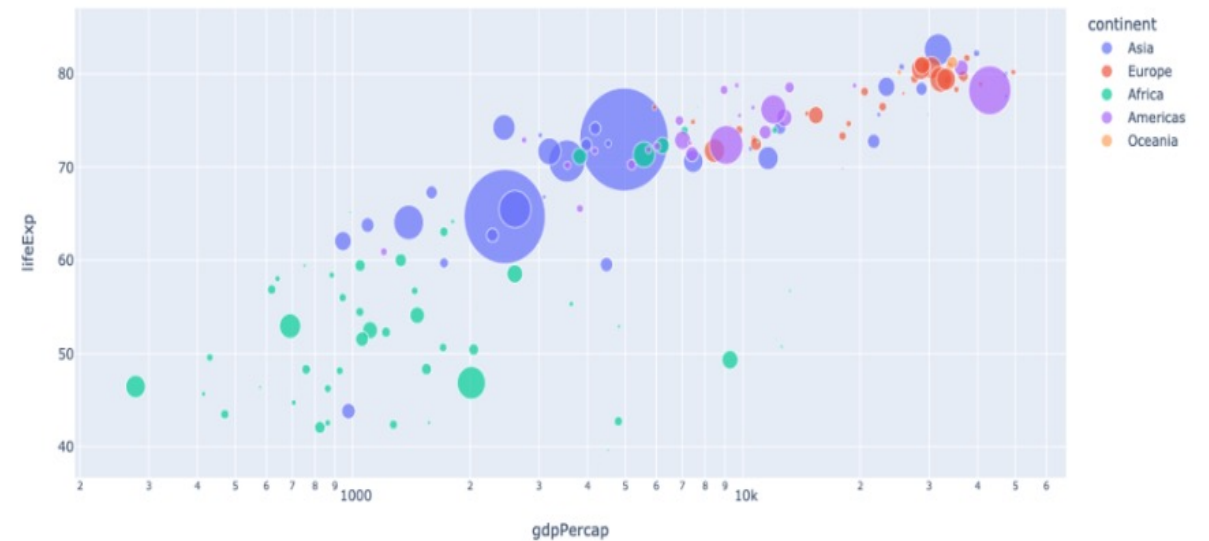
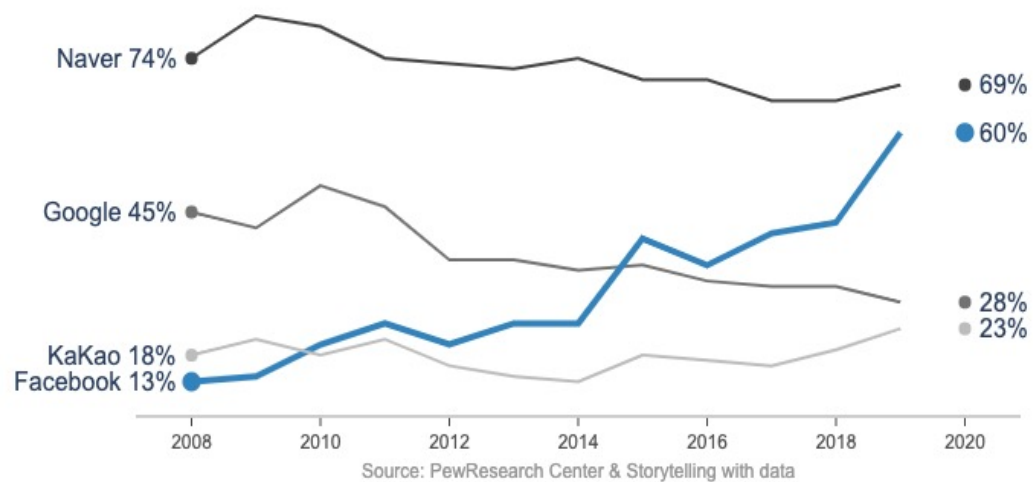


Woody K  
Contact: [woody35545@gmail.com](mailto:woody35545@gmail.com)  
Github: <https://github.com/woody35545>  
CNU, Computer Science and Engineering

# Well-Known Libraries

## Plotly

### Media Cost Percentage



Woody K  
Contact: [woody35545@gmail.com](mailto:woody35545@gmail.com)  
Github: <https://github.com/woody35545>  
CNU, Computer Science and Engineering

## 2. How to Use



# Data Visualization for Python

## 2-1. How to Use **matplotlib**



Woody K  
Contact: [woody35545@gmail.com](mailto:woody35545@gmail.com)  
Github: <https://github.com/woody35545>  
CNU, Computer Science and Engineering

# Matplotlib 사용법

## Pyplot으로 간단한 그래프 생성

`plt.plot([1, 2, 3, 4])` y-values

### 예제 1

```
import matplotlib.pyplot as plt

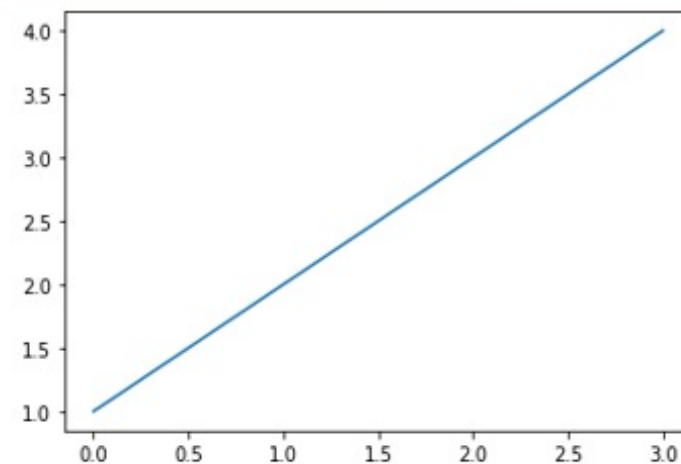
plt.plot([1, 2, 3, 4])
plt.show()
```

`pyplot.plot()` 함수에 하나의 숫자 리스트를 입력함으로써 아래와 같은 그래프가 그려집니다.

`plot()` 함수는 리스트의 값들이 y 값들이라고 가정하고, x 값 [0, 1, 2, 3]을 자동으로 만들어냅니다.

`matplotlib.pyplot` 모듈의 `show()` 함수는 그래프를 화면에 나타나도록 합니다.

생성된 그래프



Woody K  
Contact: [woody35545@gmail.com](mailto:woody35545@gmail.com)  
Github: <https://github.com/woody35545>  
CNU, Computer Science and Engineering

# Matplotlib 사용법

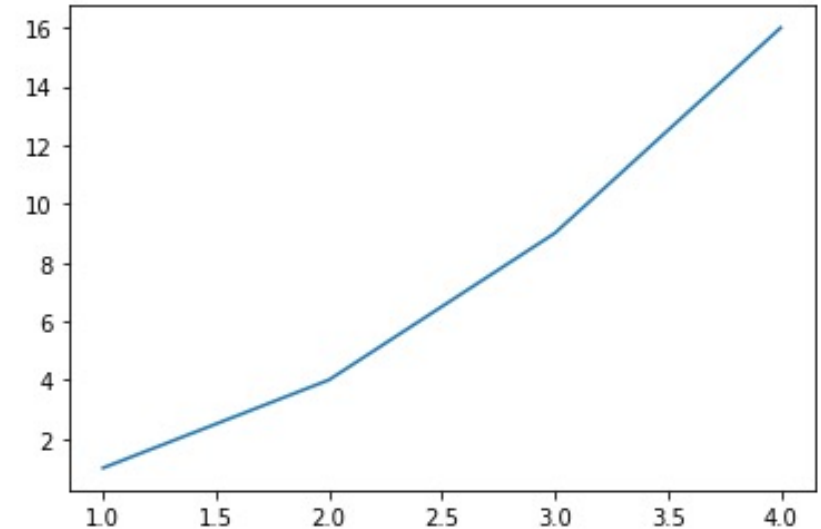
## Pyplot으로 간단한 그래프 생성

### ■ X, Y 값을 입력하여 그래프 생성

`plt.plot([ #x 좌표 ], [ # y좌표 ] )`

```
import matplotlib.pyplot as plt

plt.plot([1, 2, 3, 4], [1, 4, 9, 16])
plt.show()
```



## Format String 설정하기

Format string ↗

```
plt.plot([1, 2, 3, 4], [1, 4, 9, 16], 'ro')  
plt.axis([xmin, xmax, ymin, ymax])
```

```
import matplotlib.pyplot as plt
```

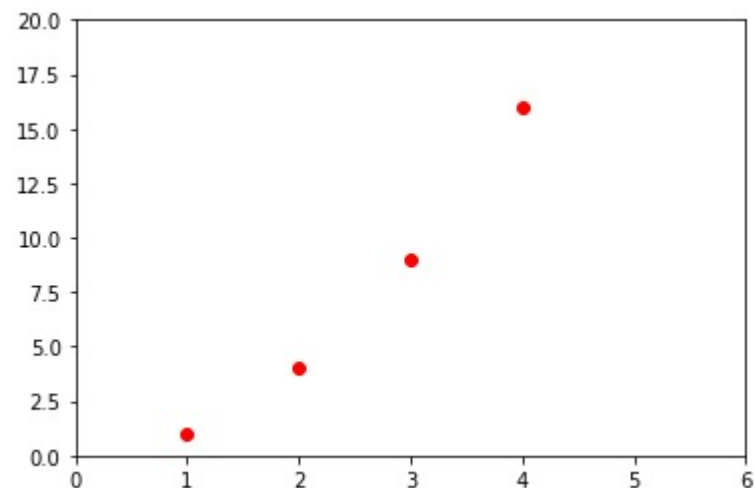
```
plt.plot([1, 2, 3, 4], [1, 4, 9, 16], 'ro')  
plt.axis([0, 6, 0, 20])  
plt.show()
```

x, y 값 인자에 대해 선의 색상과 형태를 지정하는 포맷 문자열 (**Format string**)을 세번째 인자에 입력할 수 있습니다.

포맷 문자열 'ro'는 빨간색 ('red')의 원형 ('o') 마커를 의미합니다.

또한, 예를 들어 'b-'는 파란색 ('blue')의 실선 ('-')을 의미합니다.

matplotlib.pyplot 모듈의 **axis()** 함수를 이용해서 축의 범위 [xmin, xmax, ymin, ymax]를 지정했습니다.



# Matplotlib 사용법

## Format String 설정하기

```
plt.plot(X, Y1, 'r--', X, Y2, 'bs', X, Y3, 'g^')
```

```
import matplotlib.pyplot as plt
import numpy as np

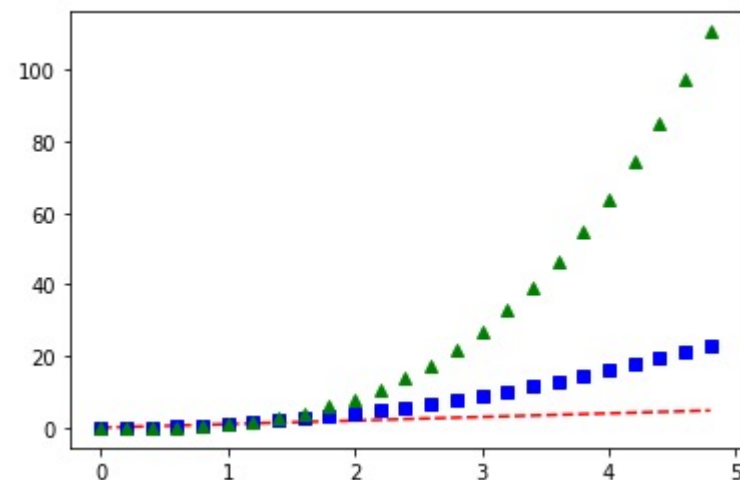
# 200ms 간격으로 균일하게 샘플된 시간
t = np.arange(0., 5., 0.2)

# 빨간 대쉬, 파란 사각형, 녹색 삼각형
plt.plot(t, t, 'r--', t, t**2, 'bs', t, t**3, 'g^')
plt.show()
```

Matplotlib에서는 일반적으로 NumPy 어레이를 이용하게 되는데,

사실 NumPy 어레이를 사용하지 않더라도 모든 시퀀스는 내부적으로 NumPy 어레이로 변환됩니다.

이 예제는 다양한 스타일을 갖는 여러 개의 곡선을 하나의 그래프로 나타냅니다.



Woody K  
Contact: [woody35545@gmail.com](mailto:woody35545@gmail.com)  
Github: <https://github.com/woody35545>  
CNU, Computer Science and Engineering

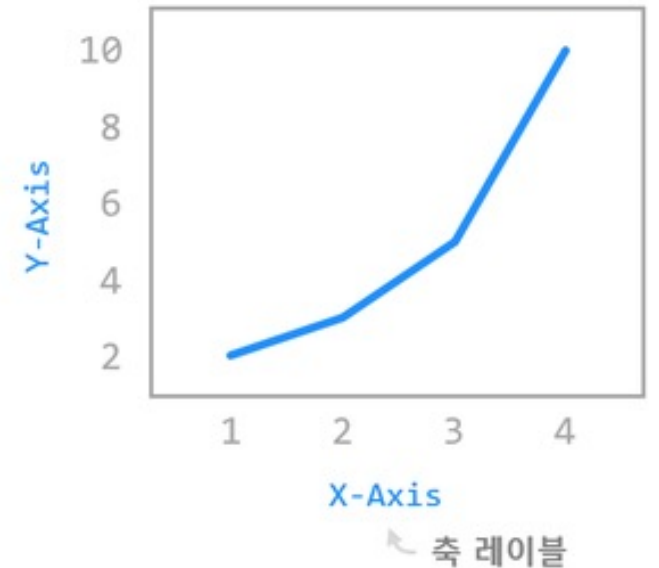
# Matplotlib 사용법

## 축 레이블 설정

`matplotlib.pyplot` 모듈의 `xlabel()`, `ylabel()` 함수를 사용하면 그래프의 x, y 축에 대한 레이블을 표시할 수 있습니다.

이 페이지에서는 `xlabel()`, `ylabel()` 함수를 사용해서 그래프의 축에 레이블을 표시하는 방법에 대해 소개합니다.

**Keyword:** `plt.xlabel()`, `plt.ylabel()`, `plt.axis()`, `labelpad`, `fontdict`, `loc`



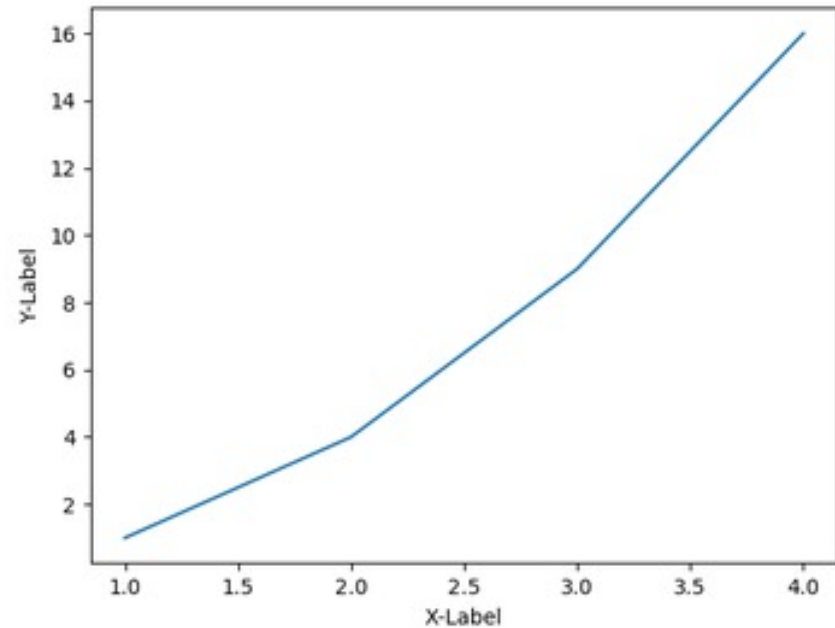
# Matplotlib 사용법

## 축 레이블 설정

### 기본적인 축 레이블 지정

```
plt.plot(x, y)  
plt.xlabel('X-Axis')  
plt.ylabel('Y-Axis')
```

```
import matplotlib.pyplot as plt  
  
plt.plot([1, 2, 3, 4], [1, 4, 9, 16])  
plt.xlabel('X-Label')  
plt.ylabel('Y-Label')  
plt.show()
```



xlabel(), ylabel() 함수에 문자열을 입력하면, 그림과 같이 각각의 축에 레이블이 표시됩니다.



# Matplotlib 사용법

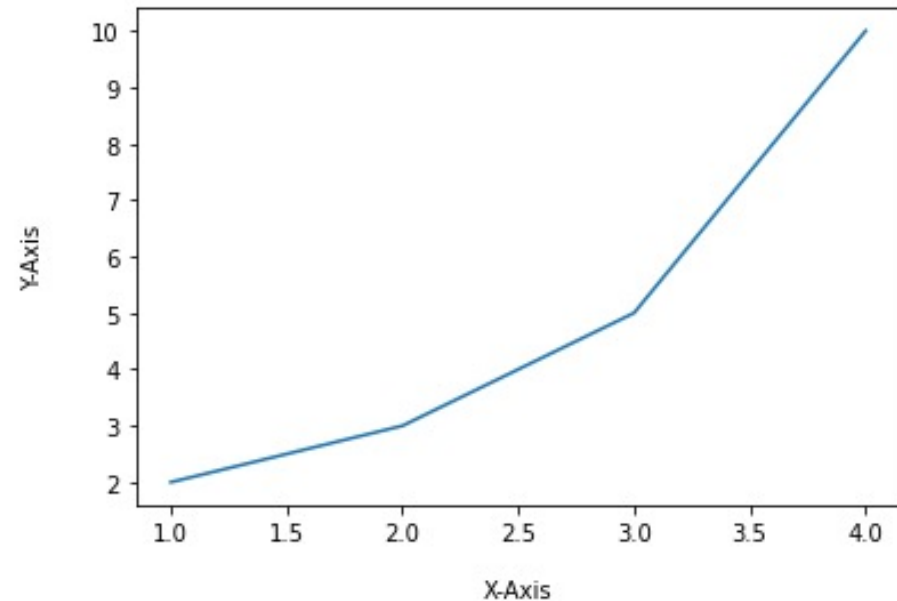
## 축 레이블 설정

### 축 레이블 여백 지정

```
plt.plot(x, y)
plt.xlabel('X-Axis', labelpad=15)
plt.ylabel('Y-Axis', labelpad=20)
```

```
import matplotlib.pyplot as plt

plt.plot([1, 2, 3, 4], [2, 3, 5, 10])
plt.xlabel('X-Axis', labelpad=15)
plt.ylabel('Y-Axis', labelpad=20)
plt.show()
```



xlabel(), ylabel() 함수의 labelpad 파라미터는 축 레이블의 **여백** (Padding)을 지정합니다.

예제에서는 x축 레이블에 대해서 15pt, y축 레이블에 대해서 20pt 만큼의 여백을 지정했습니다.



Woody K  
Contact: woody35545@gmail.com  
Github: <https://github.com/woody35545>  
CNU, Computer Science and Engineering

# Matplotlib 사용법

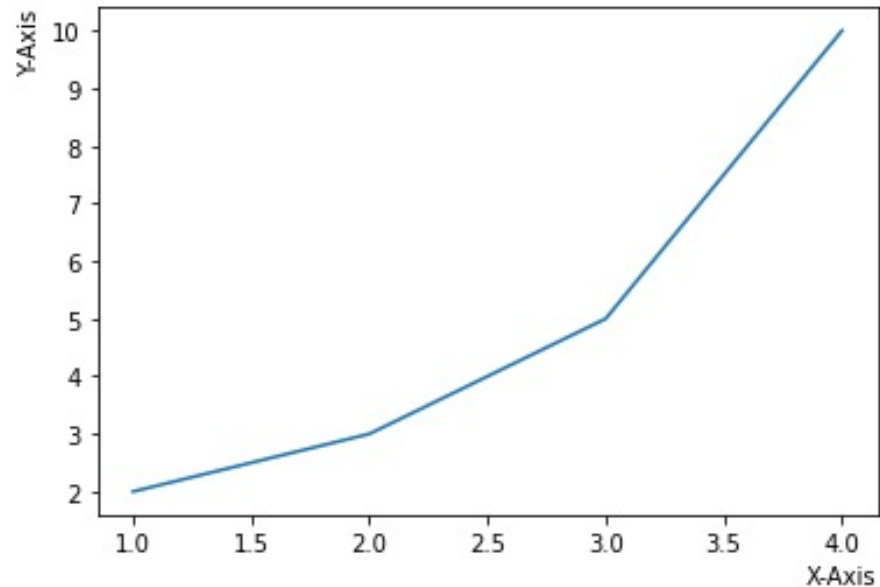
## 축 레이블 설정

### 축 레이블 위치 지정

```
plt.plot(x, y)
plt.xlabel('X-Axis', loc='right')
plt.ylabel('Y-Axis', loc='top')
```

```
import matplotlib.pyplot as plt

plt.plot([1, 2, 3, 4], [2, 3, 5, 10])
plt.xlabel('X-Axis', loc='right')
plt.ylabel('Y-Axis', loc='top')
plt.show()
```



xlabel() 함수의 loc 파라미터는 x축 레이블의 위치를 지정합니다. ({'left', 'center', 'right'})

ylabel() 함수의 loc 파라미터는 y축 레이블의 위치를 지정합니다. ({'bottom', 'center', 'top'})

이 파라미터는 Matplotlib 3.3 이후 버전부터 적용되었습니다.

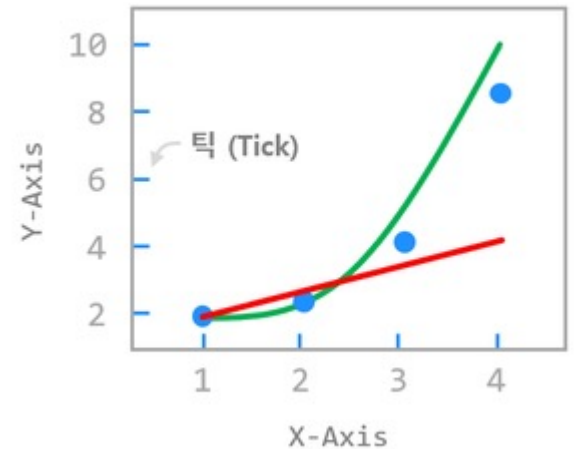


# Matplotlib 사용법

## 눈금 설정 (Ticks)

틱 (Tick)은 그래프의 축에 간격을 구분하기 위해 표시하는 눈금입니다. `matplotlib.pyplot` 모듈의 `xticks()`, `yticks()`, `tick_params()` 함수를 이용해서 그래프에 눈금을 표시하는 방법에 대해 소개합니다.

**Keyword:** `plt.xticks()`, `plt.yticks()`, `plt.tick_params()`, tick, 눈금 스타일



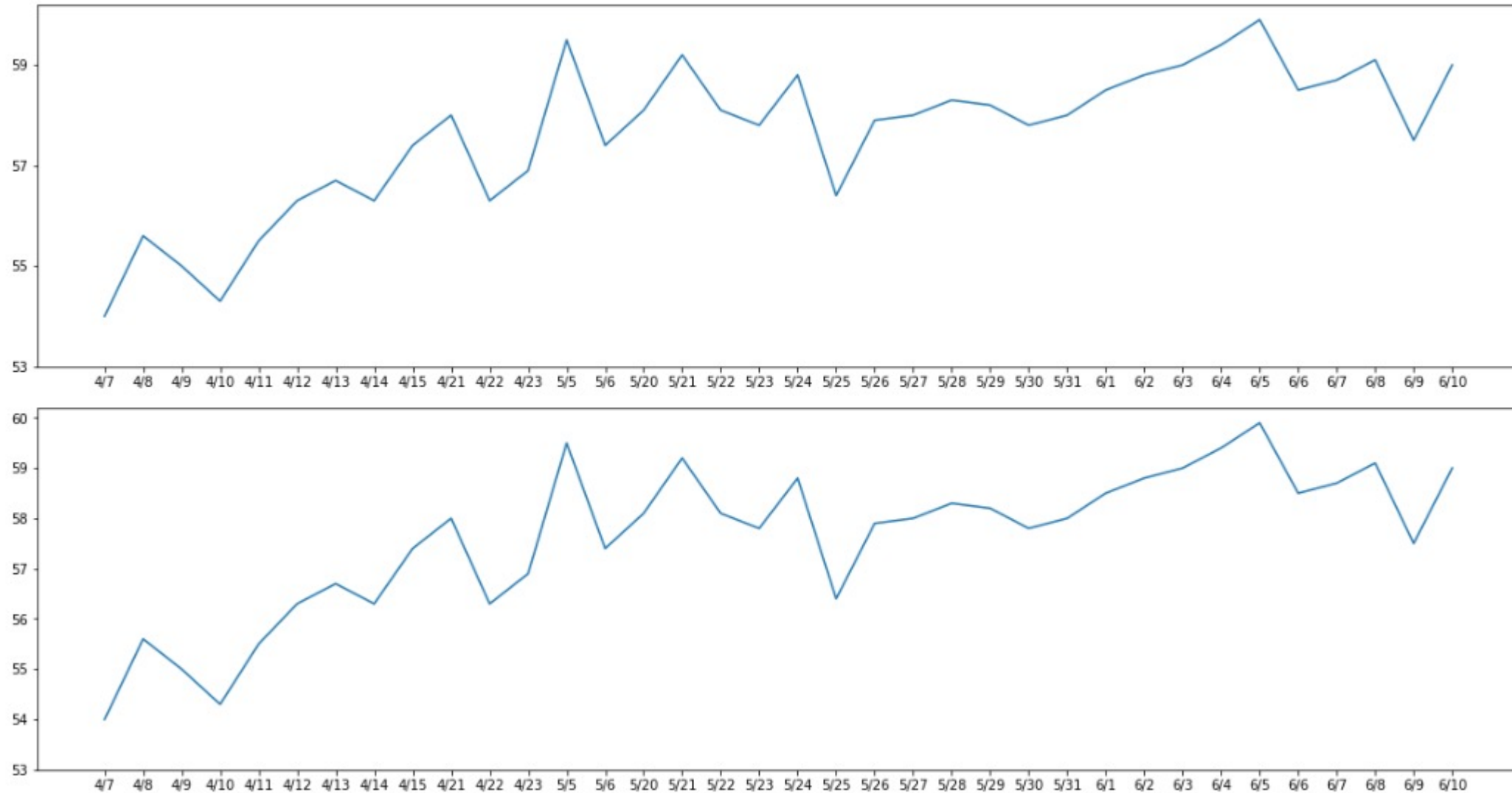
## Matplotlib Marker 종류

marker	symbol	description
"."	•	point
","	,	pixel
"o"	○	circle
"v"	▼	triangle_down
"^"	▲	triangle_up
"<"	◀	triangle_left
">"	▶	triangle_right
"1"	⋿	tri_down
"2"	⋈	tri_up
"3"	⋊	tri_left
"4"	⋋	tri_right
"8"	⬢	octagon
"s"	■	square
"p"	⬠	pentagon
"P"	⬢	plus (filled)
"*"	★	star
"h"	⬡	hexagon1
"H"	⬢	hexagon2
"+"	+	plus
"x"	×	x
"X"	⊗	x (filled)
"D"	◆	diamond
"d"	◇	thin_diamond
" "		vline
"_"	—	hline

marker	symbol	description
0 (TICKLEFT)	—	tickleft
1 (TICKRIGHT)	—	tickright
2 (TICKUP)		tickup
3 (TICKDOWN)		tickdown
4 (CARETLEFT)	◀	caretleft
5 (CARETRIGHT)	▶	caretright
6 (CARETUP)	▲	caretup
7 (CARETDOWN)	▼	caretdown
8 (CARETLEFTBASE)	◀	caretleft (centered at base)
9 (CARETRIGHTBASE)	▶	caretright (centered at base)
10 (CARETUPBASE)	▲	caretup (centered at base)
11 (CARETDOWNBASE)	▼	caretdown (centered at base)
"None", " " or ""		nothing
'\$...\$'	f	Render the string using mathtext. E.g "\$f\$" for marker showing the letter f.



## 축 레이블 눈금(ticks) 지정 예시



## References

[Matplotlib Tutorial - 파이썬으로 데이터 시각화 하기] : <https://wikidocs.net/book/5011>

Matplotlib Tutorial - 파이썬으로 데이터 시각화하기



지은이 : Dardao

최종 편집일시 : 2022년 4월 23일 2:02 오후

저작권 : (CC) BY-NC-ND

👍 113 명이 추천

[Plotly로 예쁜 그래프 그리기] :

<https://blog.naver.com/PostView.naver?blogId=shapesoflife&logNo=222448780326&parentCategoryNo=&categoryNo=&viewDate=&isShowPopularPosts=false&from=postView>



Woody K  
Contact: [woody35545@gmail.com](mailto:woody35545@gmail.com)  
Github: <https://github.com/woody35545>  
CNU, Computer Science and Engineering

# Data Visualization for Python

# END OF FILE ✓

**Woody K**

**Contact:** [woody35545@gmail.com](mailto:woody35545@gmail.com)

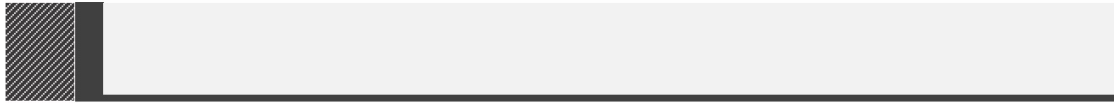
**Github:** <https://github.com/woody35545>

*CNU, Computer Science and Engineering*



## References





**Woody K**  
Contact: [woody35545@gmail.com](mailto:woody35545@gmail.com)  
Github: <https://github.com/woody35545>  
*CNU, Computer Science and Engineering*

# Main Title

## Sub Title



**Woody K**  
Contact: [woody35545@gmail.com](mailto:woody35545@gmail.com)  
Github: <https://github.com/woody35545>  
*CNU, Computer Science and Engineering*