



데이터 시각화 입문자를 위한 경험 공유

@kenu.heo



DataViz : 넷플릭스에서 통용하는 단어 **Data Visualization**

“Dataviz signifies the practice of representing data through visual and often interactive means.”

<https://www.philsimon.com/blog/data/big-data/netflix-dataviz/>



오늘 얘기는

- 머신러닝 배우다 발견한 데이터 시각화 경험



(머신러닝은 아직도 못하고 있음)

- 파이썬으로 쉽게 그리기
- 강력추천! 시각화 도구들



<http://m.blog.naver.com/proginner123/183816462>

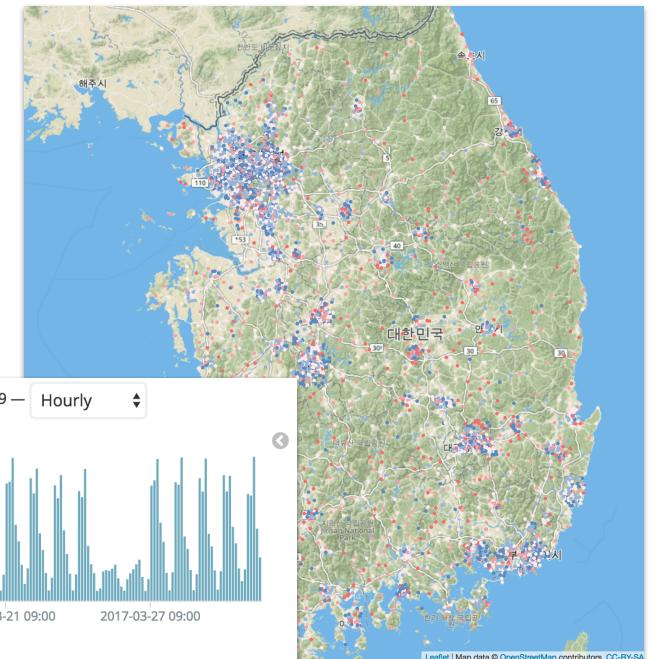
머신러닝 배우다 데이터 시각화 샛길로 빠진 얘기

2017: Tensorflow ➡ Python, jupyter ➡ Matplotlib, folium, leaflet.js

2016: Elastic Stack의 Kibana

2012~2015: node.js 의 d3.js, three.js

2000~2011: chart components





파이썬의 시각화 도구는 아름다웠습니다.

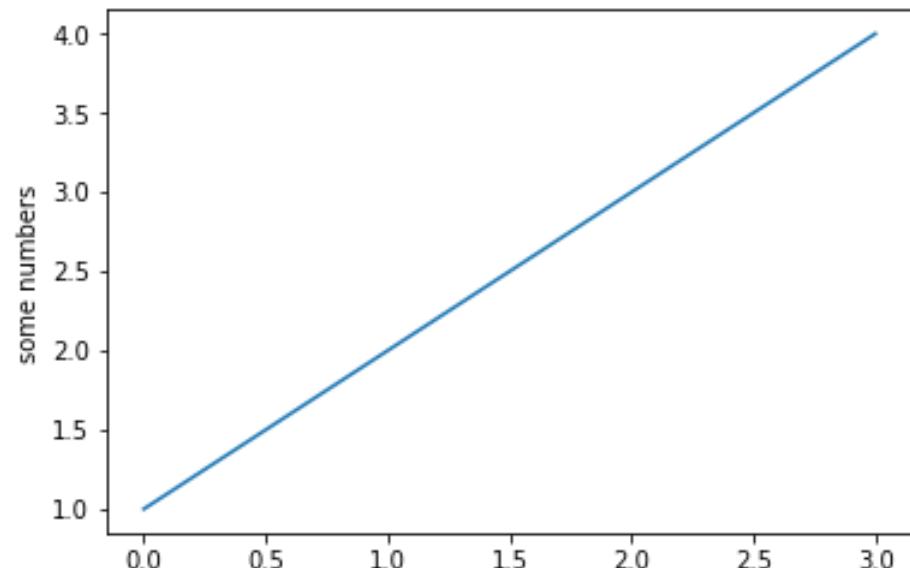
Python의 numpy, pandas로 숫자를 쉽게 다루고,
matplotlib로 쉽게 그리고,
지도가 필요하다면 folium을 이용하면 끝

pip install jupyter matplotlib pandas



파이썬으로 쉽게 그리기

```
import matplotlib.pyplot as plt  
plt.plot([1,2,3,4])  
plt.ylabel('some numbers')  
plt.show()
```



강력추천! 시각화 도구들과 정보

찾는 게 제일 힘들었어요.

불친절한 구글
안알랴줌

그래서, 공유합니다.

지도라면 leaflet, folium



an open-source JavaScript library
for mobile-friendly interactive maps

엑셀, csv는 pandas + matplotlib

```
In [10]: import pandas as pd  
data = pd.read_csv('data/bar.csv')  
data
```

Out[10]:

	@timestamp per day	Count
0	1489590000000	174963
1	1489676400000	157612
2	1489762800000	63738
3	1489849200000	76471
4	1489935600000	185640
5	1490022000000	191737
6	1490108400000	185764
7	1490194800000	1260

엑셀, csv는 pandas + matplotlib

7	1490194800000	1260
---	---------------	------

```
In [11]: data['Count']
```

```
Out[11]: 0    174963
         1    157612
         2     63738
         3     76471
         4    185640
         5    191737
         6    185764
         7     1260
Name: Count, dtype: int64
```

Thanks

Kenu

kenu.heo@gmail.com

edited by @hailey