

IPA 주관 인공지능센터 기본(fundamental) 과정

- GitHub link: [here](#)
- E-Mail: windkyle7@gmail.com

Seaborn

- 공식 홈페이지: <https://seaborn.pydata.org/index.html>
- 특징
 - Made with matplotlib
 - Easy coding when using pandas
 - Various color theme
 - Various statistical graphs

```
In [1]: import matplotlib
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

현재 실습에 사용하는 matplotlib의 버전은 다음과 같다.

```
In [2]: matplotlib.__version__
```

```
Out[2]: '3.0.3'
```

tidy-data: iris 불러오기

```
In [3]: iris = sns.load_dataset('iris')
iris.sample(5)
```

```
Out[3]:
```

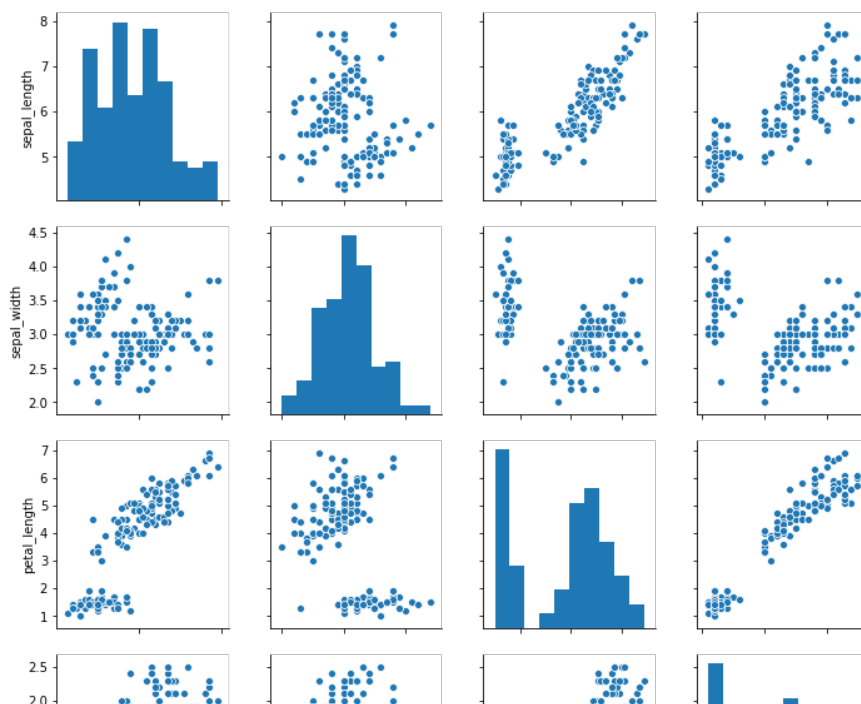
	sepal_length	sepal_width	petal_length	petal_width	species
103	6.3	2.9	5.6	1.8	virginica
17	5.1	3.5	1.4	0.3	setosa
32	5.2	4.1	1.5	0.1	setosa
26	5.0	3.4	1.6	0.4	setosa
77	6.7	3.0	5.0	1.7	versicolor

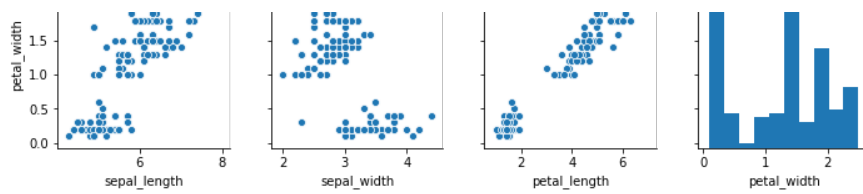
Pair plot

- 모든 변수간 scatter plot을 나타냄
- 대각선은 변수 하나의 box plot
- 적당한 변수들 상관관계 분석 가능
- PairGrid과 동일하지만 보다 쉽게 그릴 수 있음

```
In [4]: sns.pairplot(iris)
```

```
Out[4]: <seaborn.axisgrid.PairGrid at 0x7f63619add68>
```



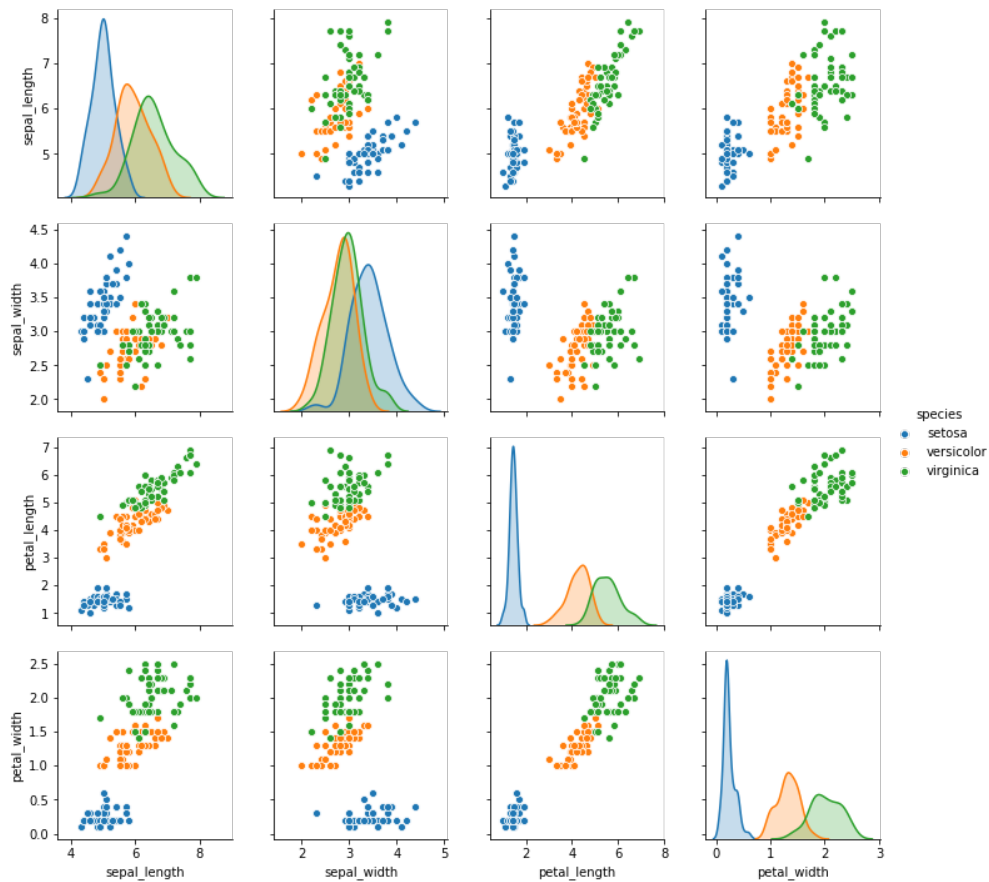


pairplot 파라미터

- hue: plot 축면을 서로 다른 색상에 매핑하기 위한 데이터 변수
- kind: 정체성이 없는 관계를 위한 plot

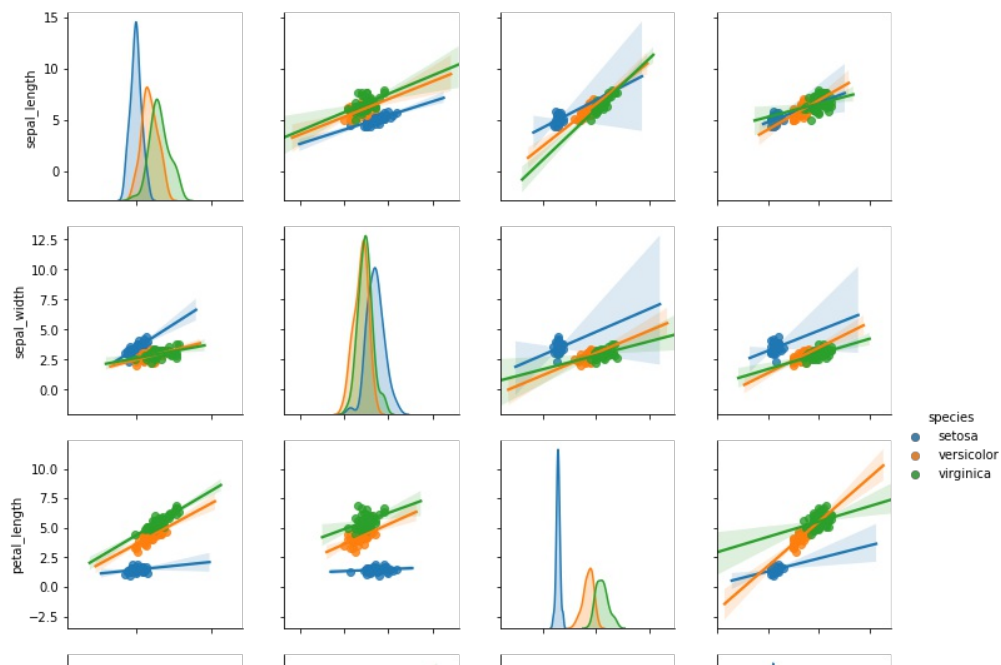
```
In [5]: sns.pairplot(iris, hue='species')
```

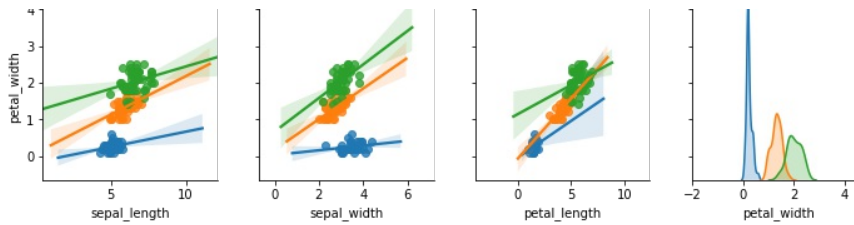
```
Out[5]: <seaborn.axisgrid.PairGrid at 0x7f635f256cc0>
```



```
In [6]: sns.pairplot(iris, hue='species', kind='reg')
```

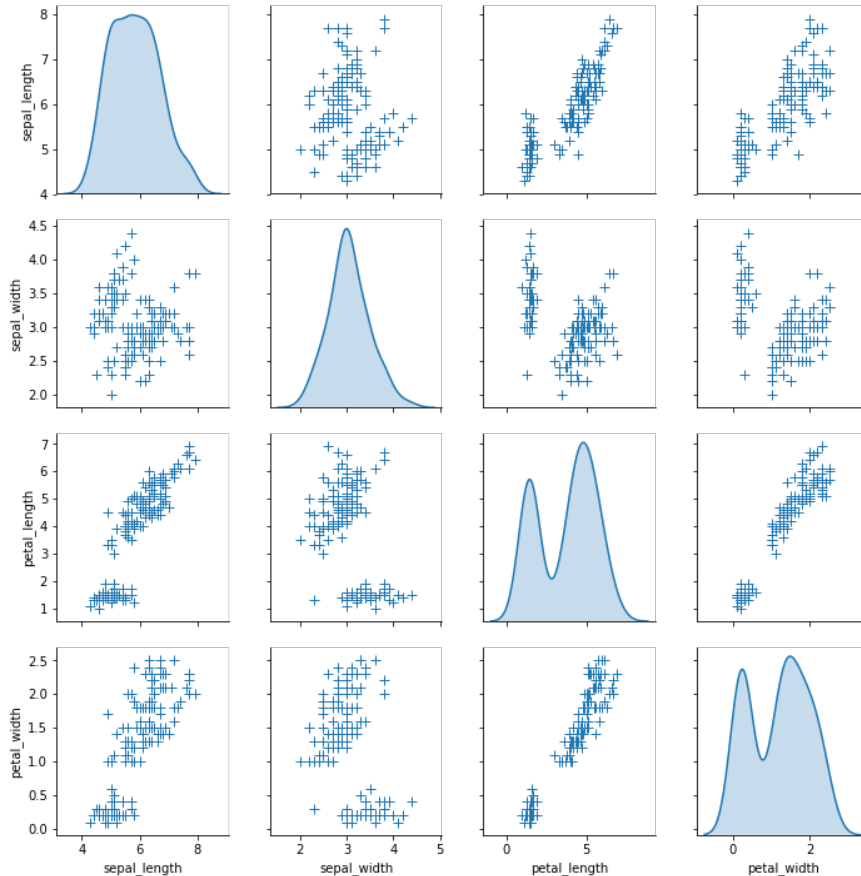
```
Out[6]: <seaborn.axisgrid.PairGrid at 0x7f63ac5d8898>
```





```
In [7]: sns.pairplot(iris,
    diag_kind='kde',
    markers='+',
    plot_kws=dict(s=50, edgecolor='b', linewidth=1),
    diag_kws=dict(shade=True))
```

Out[7]: <seaborn.axisgrid.PairGrid at 0x7f635cc09390>



tidy-data: tips 불러오기

```
In [8]: tips = sns.load_dataset('tips')
tips.sample(5)
```

Out[8]:

	total_bill	tip	sex	smoker	day	time	size
224	13.42	1.58	Male	Yes	Fri	Lunch	2
94	22.75	3.25	Female	No	Fri	Dinner	2
233	10.77	1.47	Male	No	Sat	Dinner	2
119	24.08	2.92	Female	No	Thur	Lunch	4
148	9.78	1.73	Male	No	Thur	Lunch	2

FacetGrid

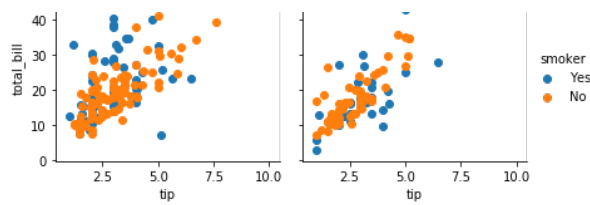
- 조건부 관계를 표시하기 위한 멀티 plot
- 데이터셋의 변수 level에 해당하는 행 및 열의 격자로 열거된 여러 축으로 데이터셋을 매핑

성별 기준으로 그리기

```
In [9]: g = sns.FacetGrid(tips, col='sex', hue='smoker')
g.map(plt.scatter, 'tip', 'total_bill')
g.add_legend()
```

Out[9]: <seaborn.axisgrid.FacetGrid at 0x7f635c14e7b8>

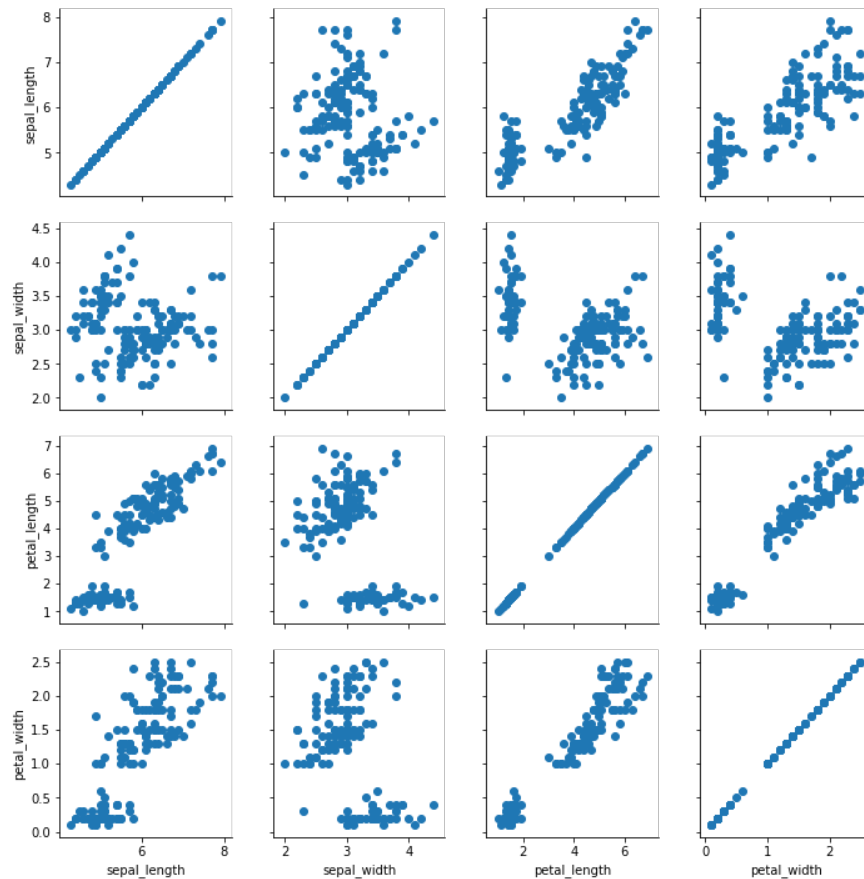




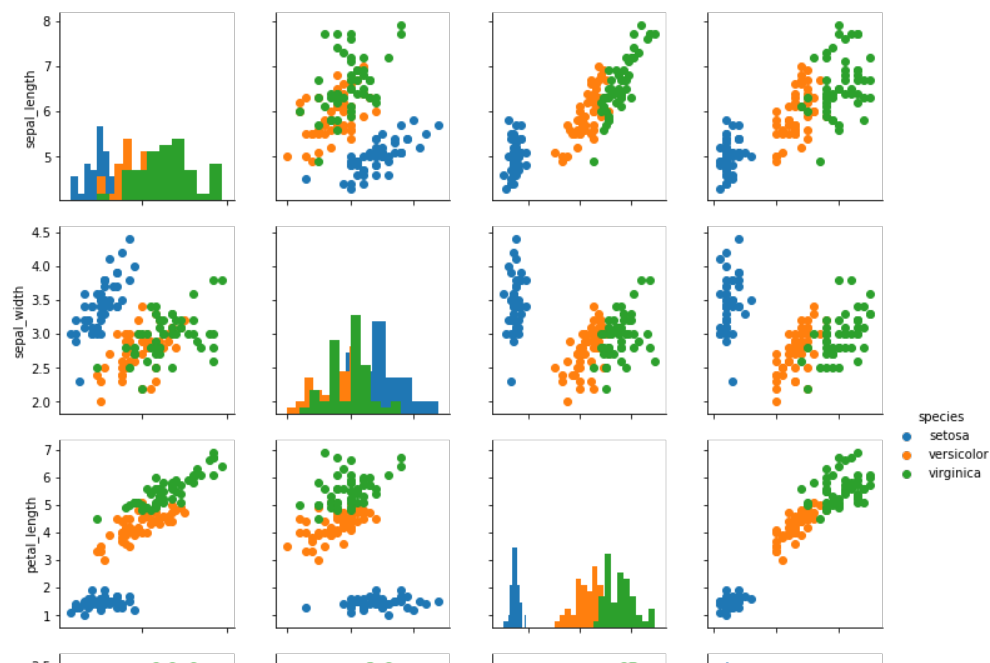
PairGrid

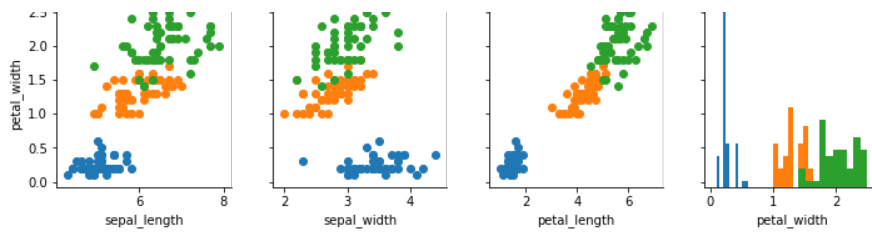
- 데이터셋의 각 변수를 여러 축의 그리드에 있는 열과 행에 매핑

```
In [10]: g = sns.PairGrid(iris)
g = g.map(plt.scatter)
```

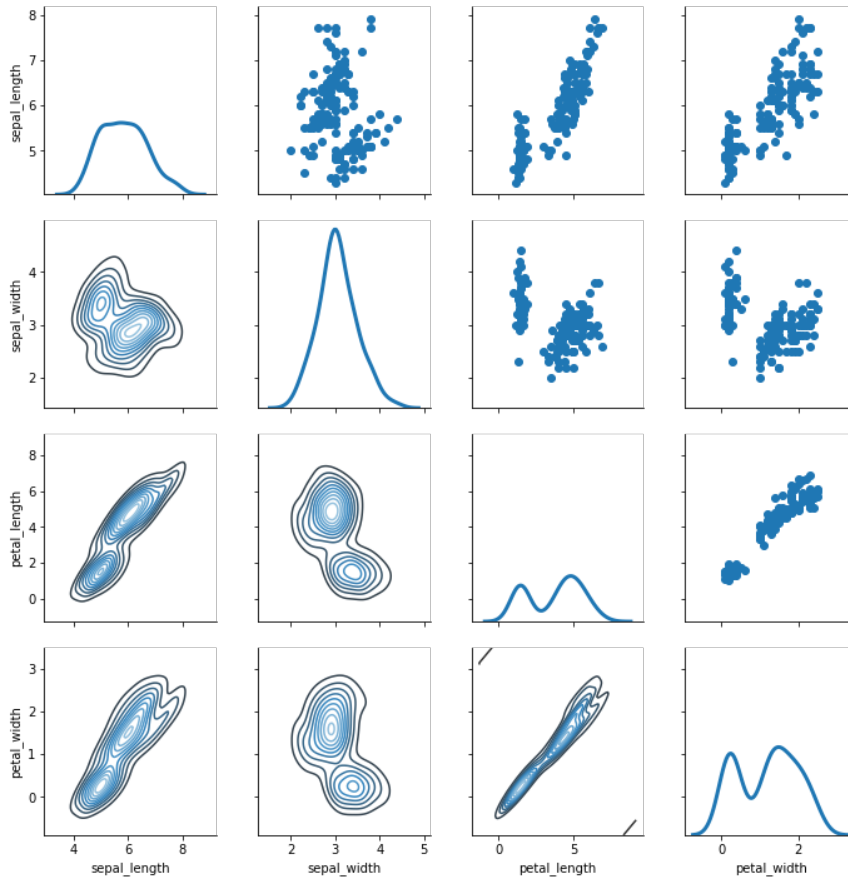


```
In [11]: g = sns.PairGrid(iris, hue='species')
g = g.map_diag(plt.hist)
g = g.map_offdiag(plt.scatter)
g = g.add_legend()
```





```
In [12]: g = sns.PairGrid(iris)
g = g.map_upper(plt.scatter)
g = g.map_lower(sns.kdeplot, cmap='Blues_d')
g = g.map_diag(sns.kdeplot, lw=3, legend=False)
```

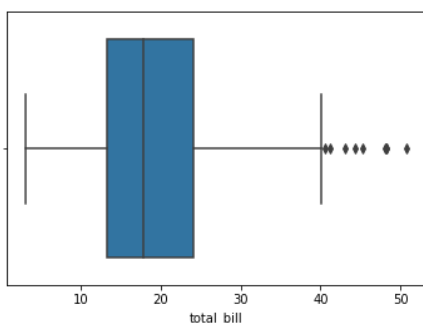


boxplot

- boxplot은 변수간 또는 범주형 변수의 레벨간 비교를 용이하게 하는 방식으로 양적 데이터의 분포를 보여줌

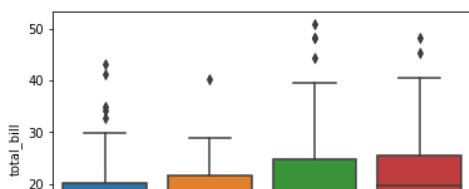
```
In [13]: sns.boxplot(x=tips['total_bill'])
```

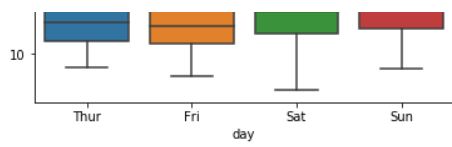
```
Out[13]: <matplotlib.axes._subplots.AxesSubplot at 0x7f6356a740f0>
```



```
In [14]: sns.boxplot(x='day', y='total_bill', data=tips)
```

```
Out[14]: <matplotlib.axes._subplots.AxesSubplot at 0x7f635692ca58>
```



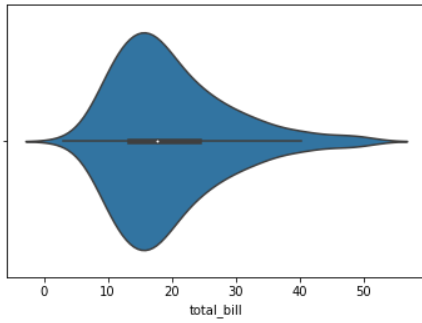


violinplot

- boxplot과 유사한 역할을 한다.

```
In [15]: sns.violinplot(x=tips['total_bill'])
```

```
Out[15]: <matplotlib.axes._subplots.AxesSubplot at 0x7f6355cd89e8>
```



```
In [16]: sns.violinplot(x='day', y='total_bill', data=tips)
```

```
Out[16]: <matplotlib.axes._subplots.AxesSubplot at 0x7f6355cad18>
```

