Seaborn绘图示例

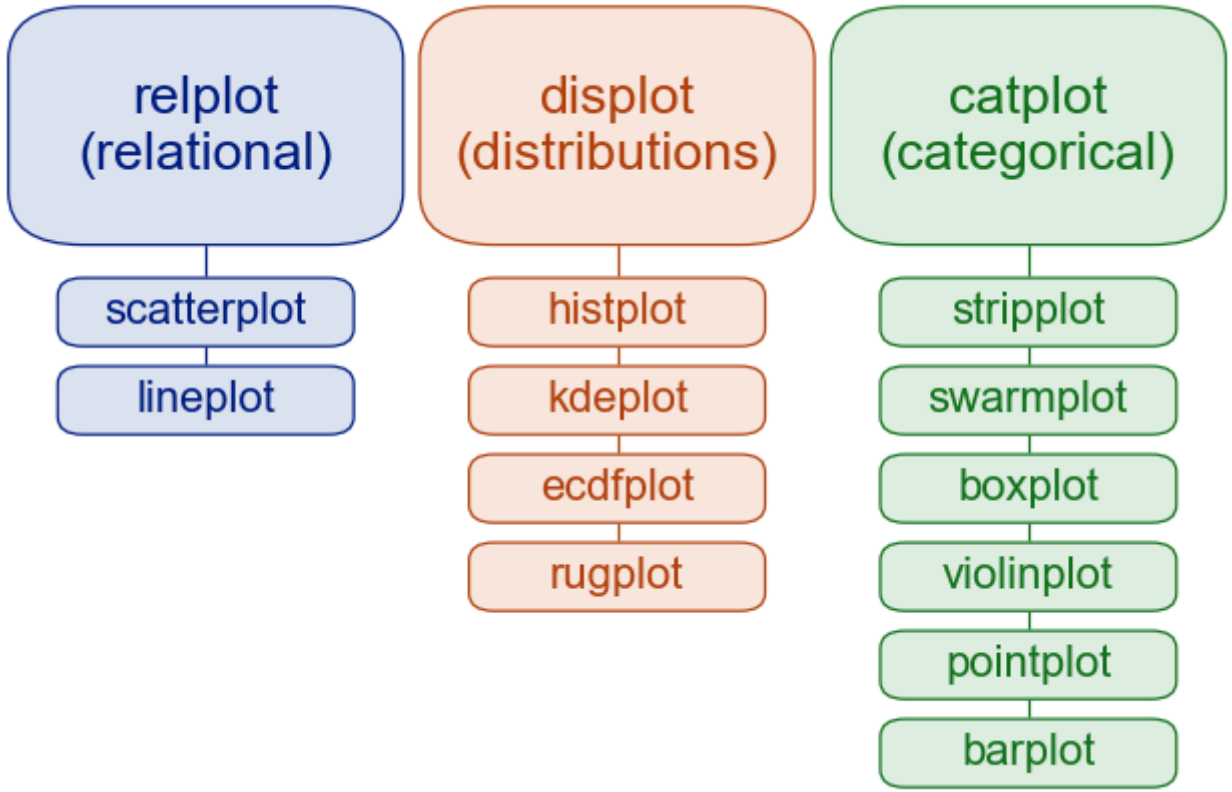
seaborn

- 安装：pip install seaborn

- seaborn是基于matplotlib的图形可视化工具
- 更高层级的封装->简洁
- matplotlib更灵活

seaborn

- figure-level vs axes-level



数据准备

```
import seaborn as sns
tips = sns.load_dataset("tips")
tips
```

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4
...	...	...	...	...	...	...	...
239	29.03	5.92	Male	No	Sat	Dinner	3
240	27.18	2.00	Female	Yes	Sat	Dinner	2
241	22.67	2.00	Male	Yes	Sat	Dinner	2
242	17.82	1.75	Male	No	Sat	Dinner	2
243	18.78	3.00	Female	No	Thur	Dinner	2

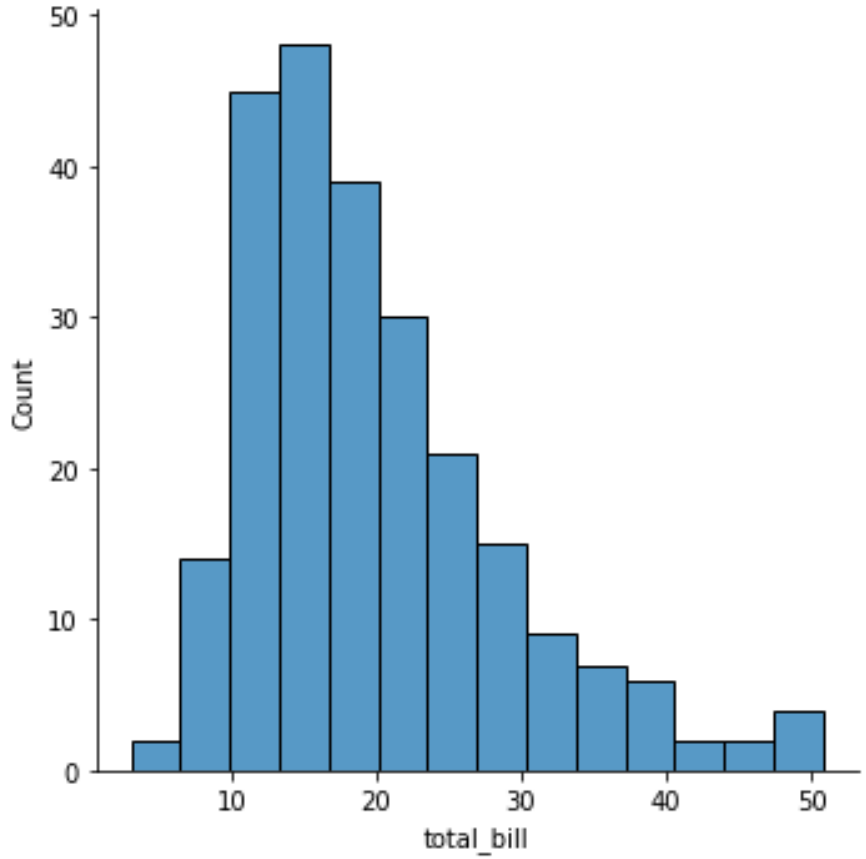
```
names = sns.get_dataset_names()
names
```

```
['anagrams',
 'anscombe',
 'attention',
 'brain_networks',
 'car_crashes',
 'diamonds',
 'dots',
 'exercise',
 'flights',
 'fmri',
 'gammas',
 'geyser',
 'iris',
 'mpg',
 'penguins',
 'planets',
 'tips',
 'titanic']
```

displot –histplot 直方图

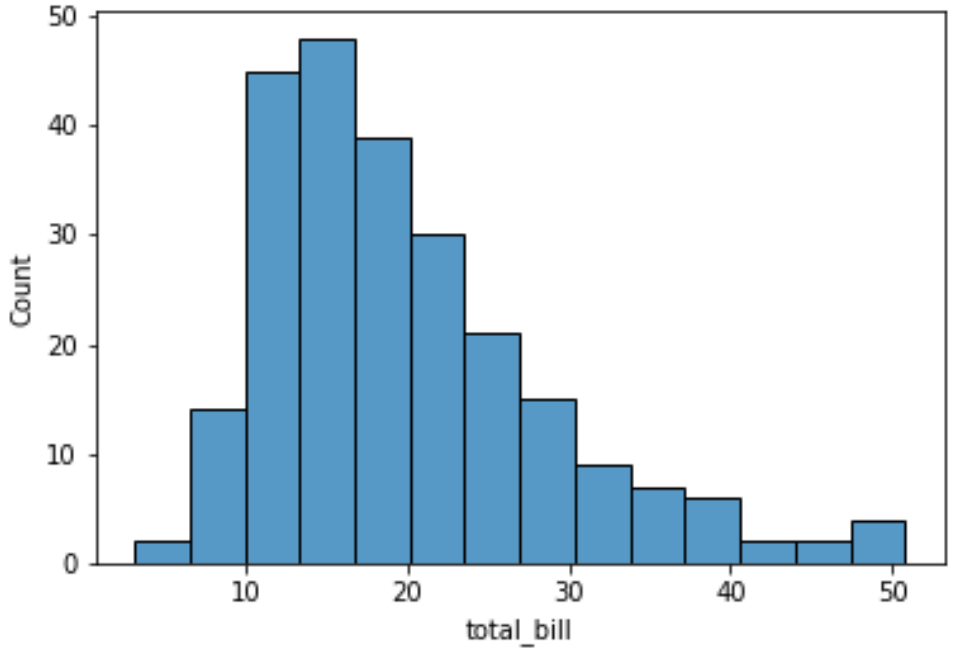
```
sns.displot(tips, x="total_bill" )
```

```
<seaborn.axisgrid.FacetGrid at 0x1dc8e65ba88>
```




```
sns.histplot(tips, x="total_bill")
```

```
<AxesSubplot:xlabel='total_bill', ylabel='Count'>
```

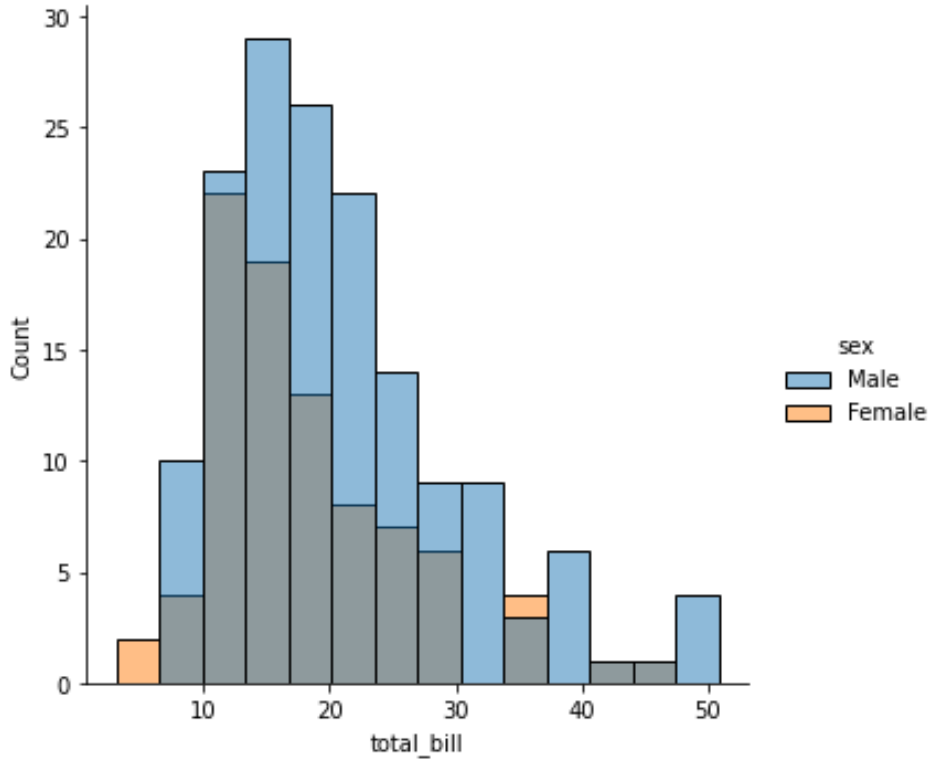


displot-分类

“hue” 参数使用不同颜色区分类别

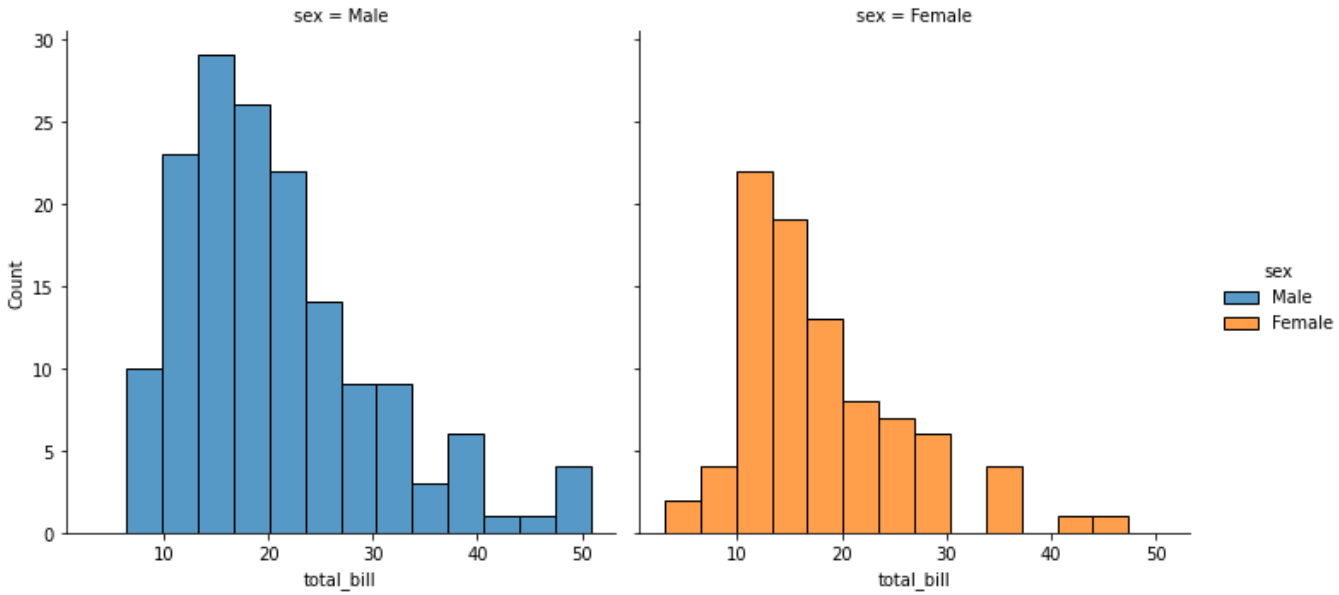
```
sns.displot(tips, x="total_bill", hue="sex")
```

```
<seaborn.axisgrid.FacetGrid at 0x1dc8e9181c8>
```




```
sns.displot(tips, x="total_bill", hue="sex",col='sex')
```

```
<seaborn.axisgrid.FacetGrid at 0x1dc8efabf48>
```

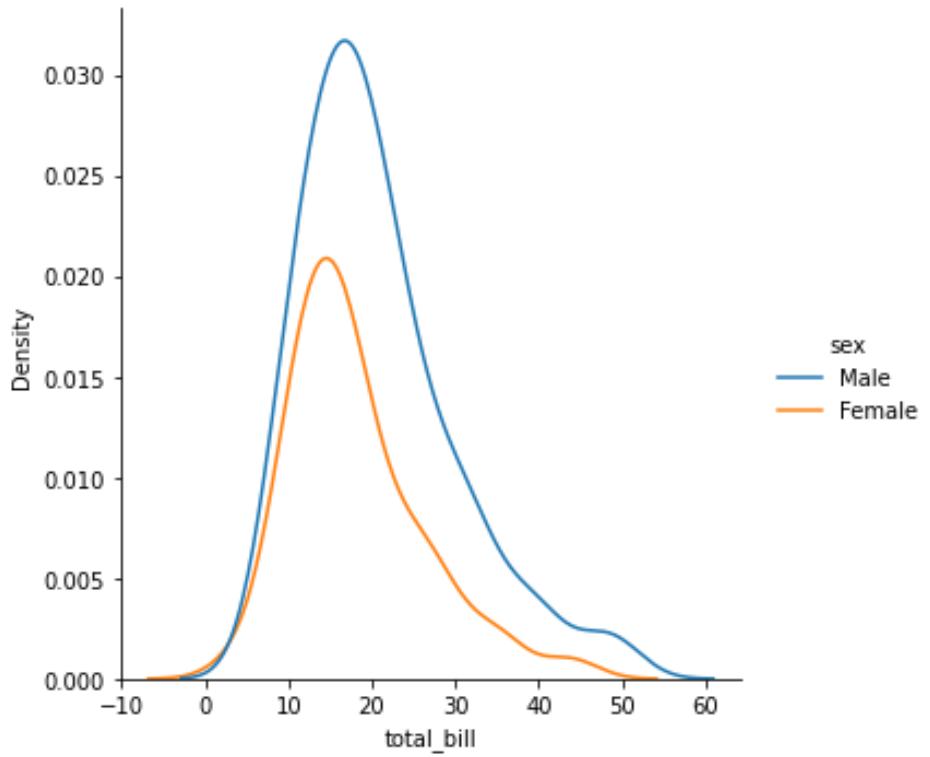


displot-kdeplot

```
sns.kdeplot(data=tips, x="total_bill",hue='sex')
```

```
sns.displot(tips, x="total_bill", kind="kde",hue='sex')
```

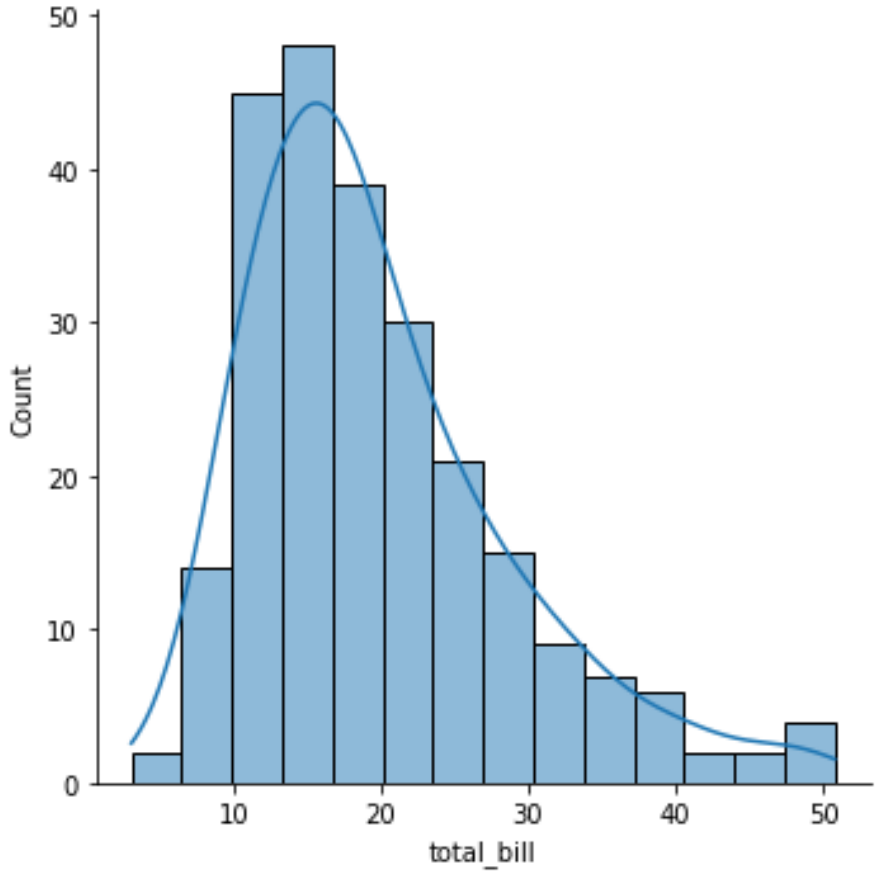
```
<seaborn.axisgrid.FacetGrid at 0x1dc8ecbe448>
```



也可同时绘制直方图和概率密度图

```
sns.displot(tips, x="total_bill", kde=True)
```

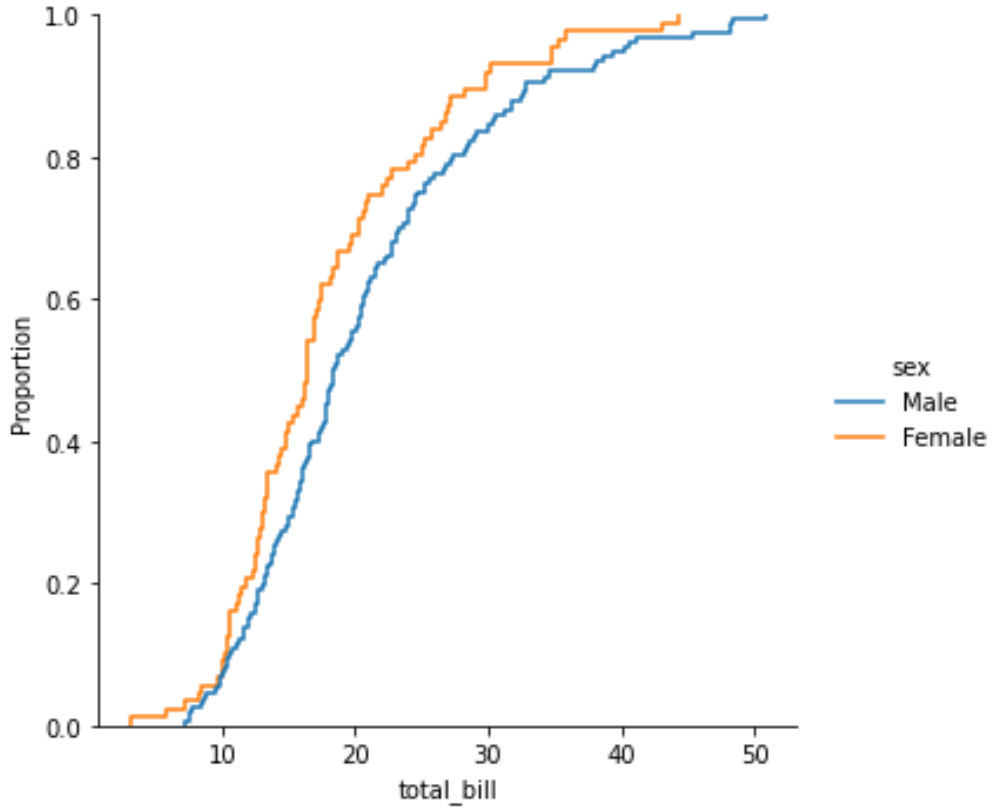
```
<seaborn.axisgrid.FacetGrid at 0x20c2d9a0c88>
```



displot-ecdf

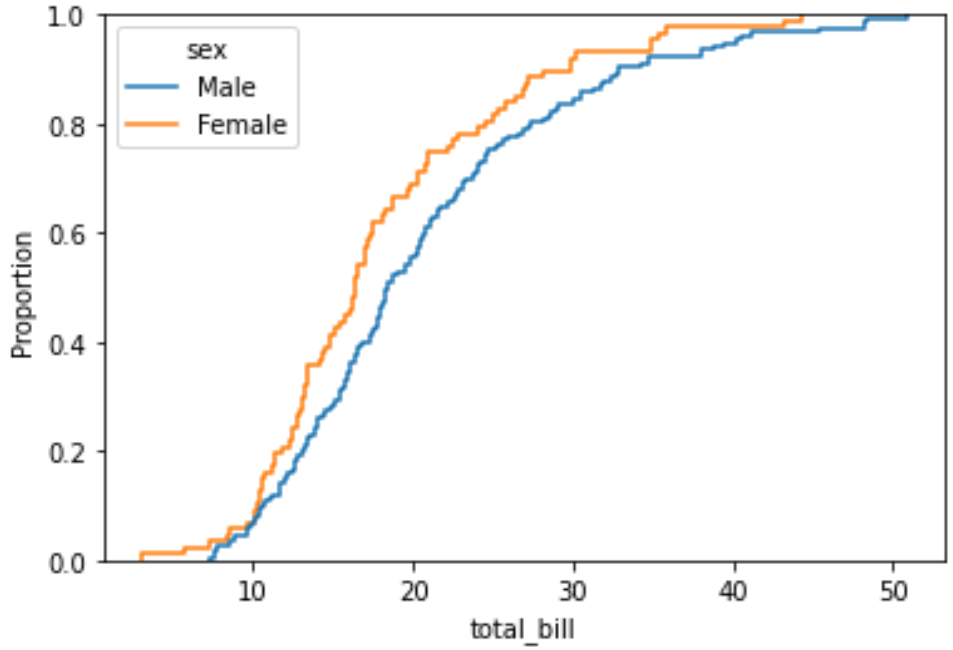
```
sns.displot(tips, x="total_bill", hue="sex", kind="ecdf")
```

```
<seaborn.axisgrid.FacetGrid at 0x1dc8f145988>
```




```
sns.ecdfplot(tips, x="total_bill", hue="sex")
```

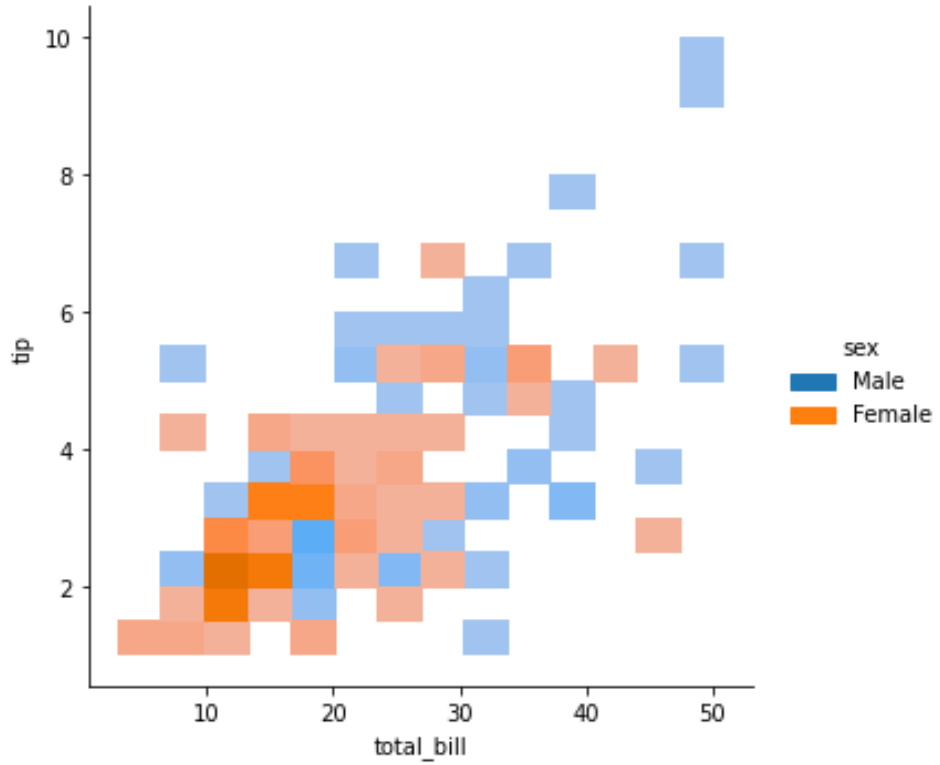
```
<AxesSubplot:xlabel='total_bill', ylabel='Proportion'>
```



displot-多变量

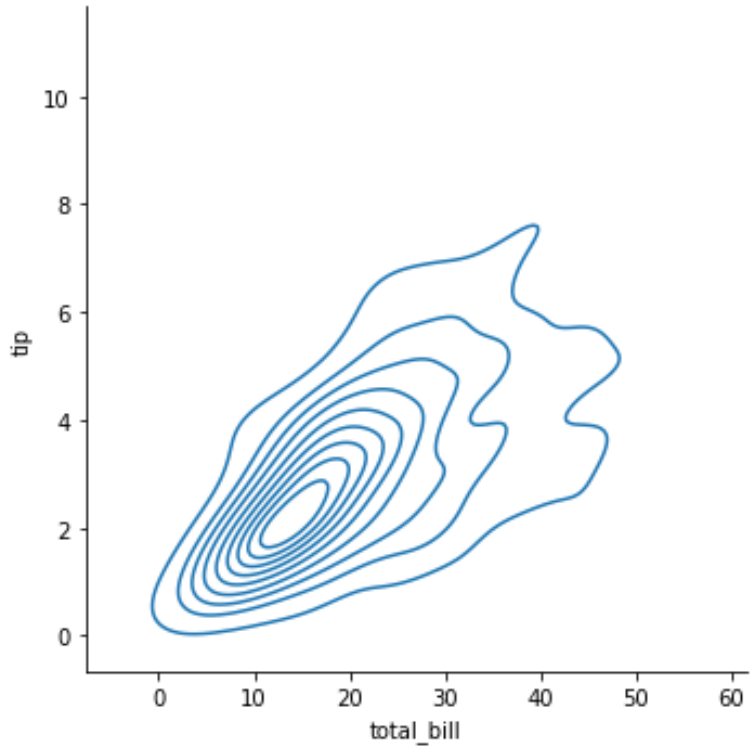
```
sns.displot(tips, x="total_bill", y="tip",hue='sex')
```

```
<seaborn.axisgrid.FacetGrid at 0x1dc8f43ba88>
```




```
sns.displot(tips, x="total_bill", y="tip", kind="kde")
```

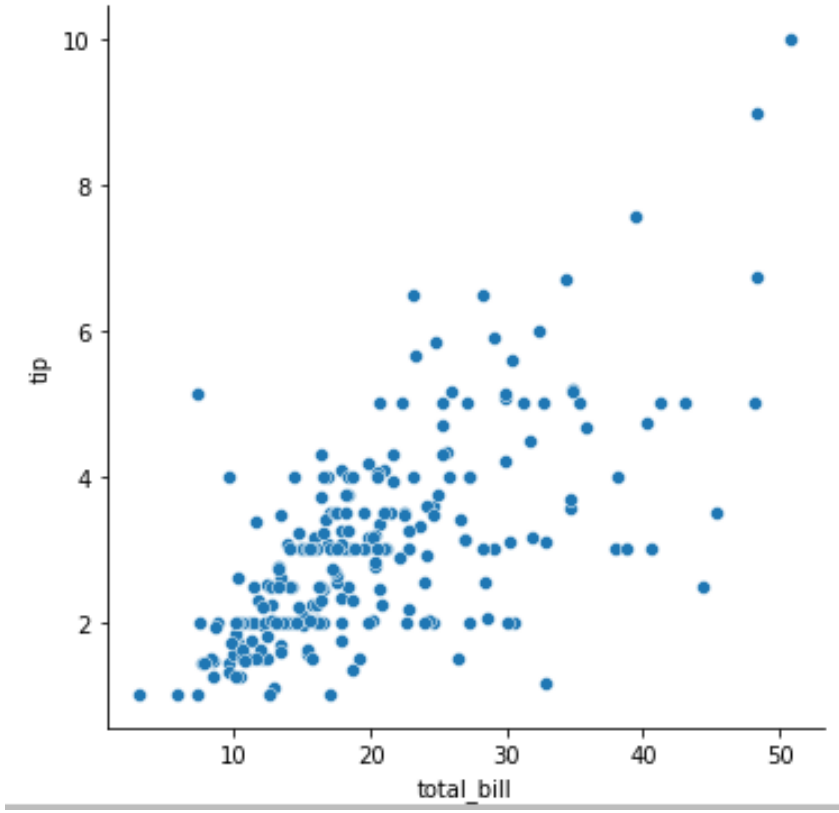
```
<seaborn.axisgrid.FacetGrid at 0x1dc8f4500c8>
```



relplot-scatterplot

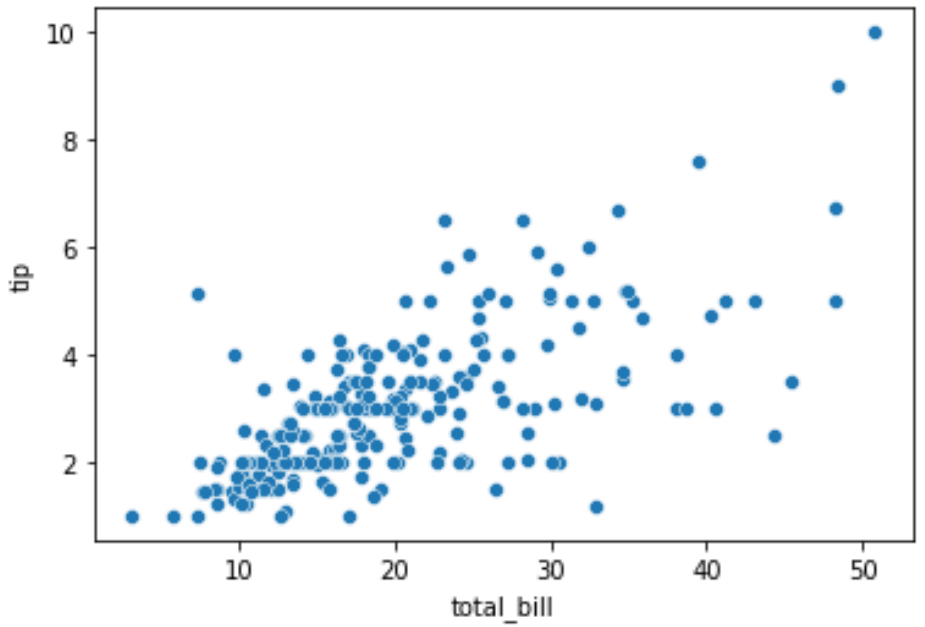
```
sns.relplot(x="total_bill", y="tip", data=tips)
```

```
<seaborn.axisgrid.FacetGrid at 0x1dc8f440508>
```




```
sns.scatterplot(x="total_bill", y="tip", data=tips)
```

```
<AxesSubplot:xlabel='total_bill', ylabel='tip'>
```

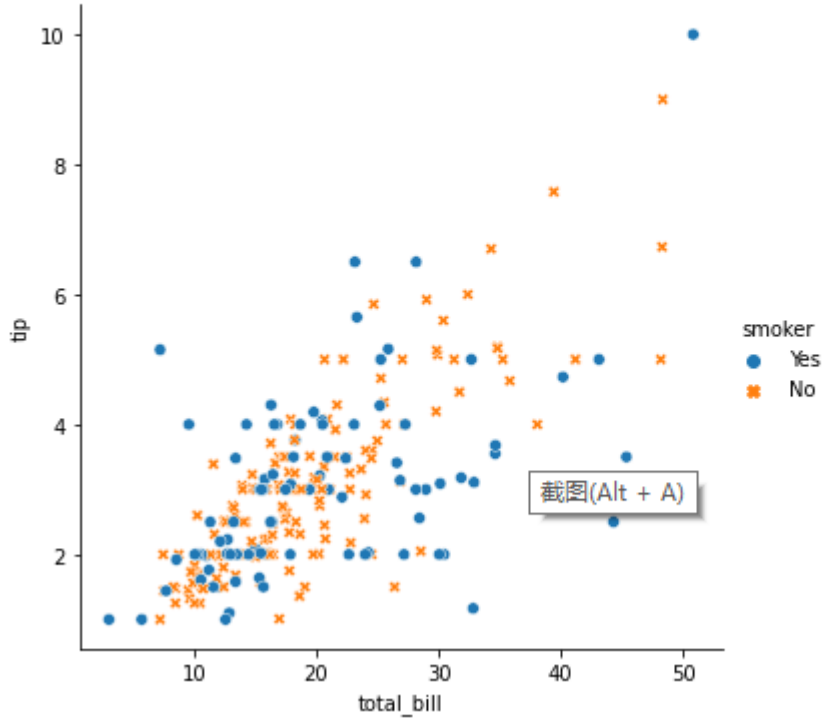


relplot-scatterplot

style参数根据类别使用不同形状，推荐和hue使用相同变量

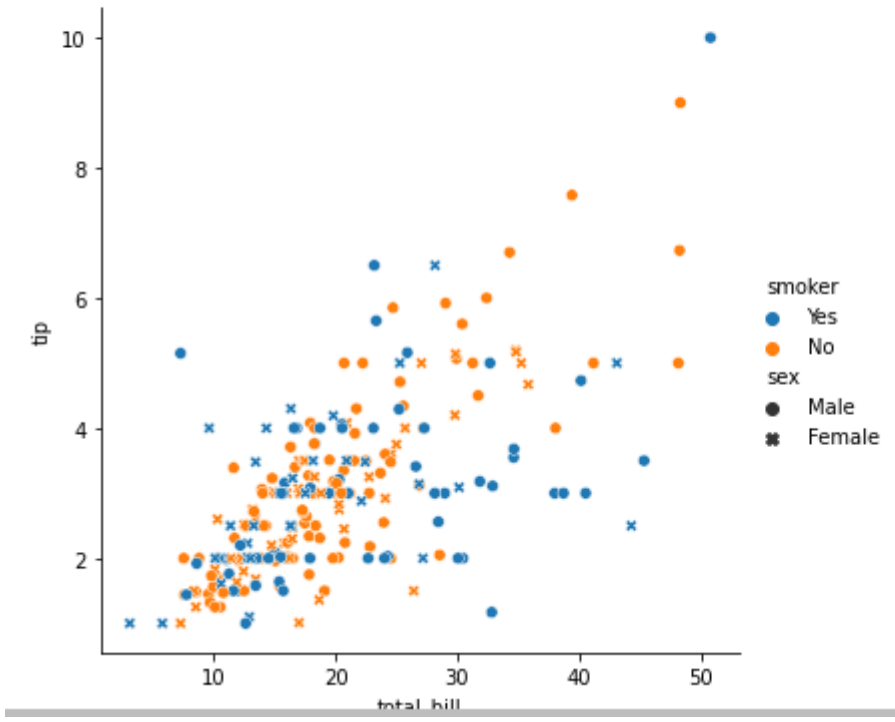
```
sns.relplot( data=tips,x="total_bill", y="tip", hue="smoker", style="smoker")
```

```
<seaborn.axisgrid.FacetGrid at 0x1dc8f517a48>
```




```
sns.relplot( data=tips,x="total_bill", y="tip", hue="smoker", style="sex")
```

```
<seaborn.axisgrid.FacetGrid at 0x20c2e313648>
```

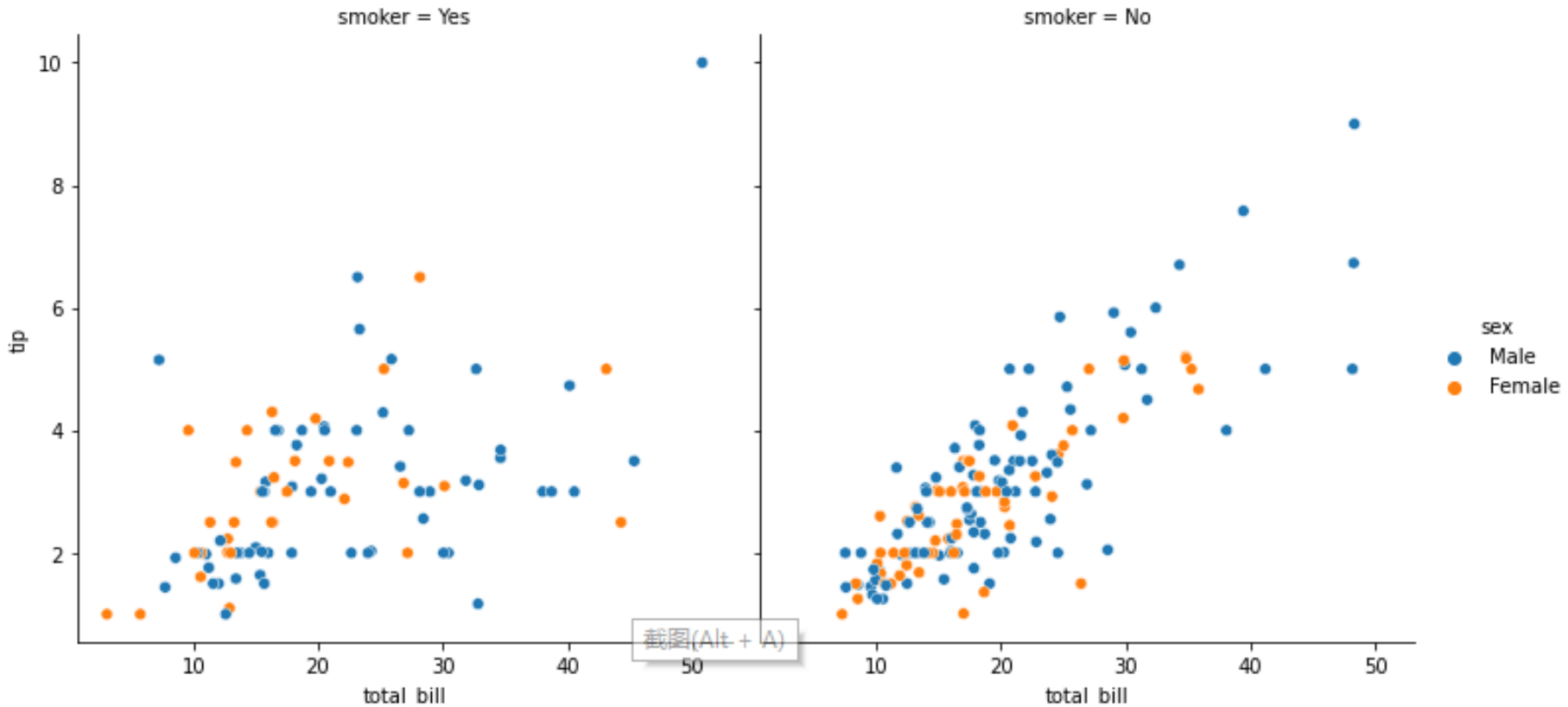


relplot

使用col或row绘制子图区分新的种类

```
sns.relplot(data=tips,x="total_bill", y="tip", hue="sex",col='smoker')
```

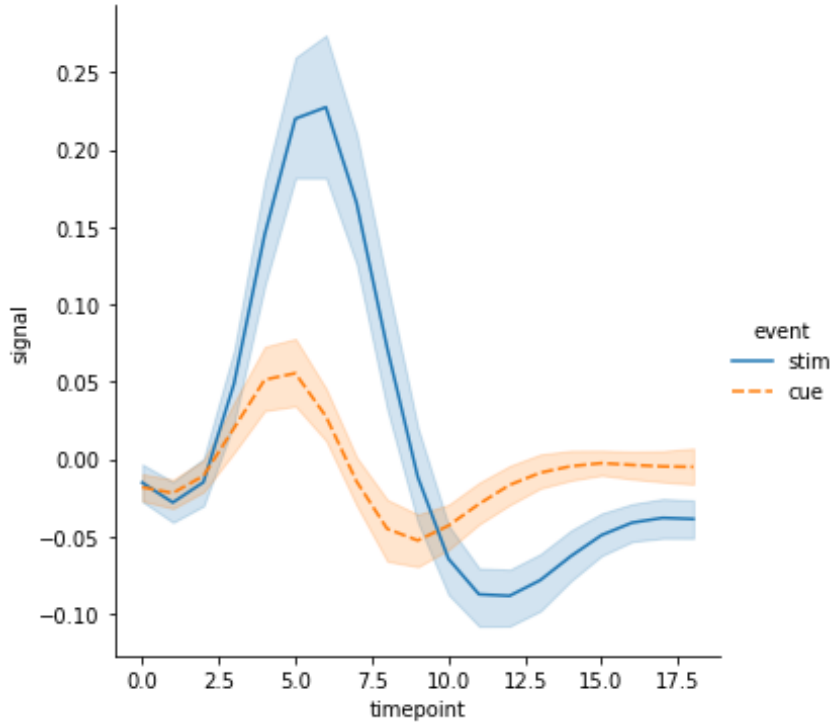
```
<seaborn.axisgrid.FacetGrid at 0x1dc93d08d08>
```



relplot-lineplot

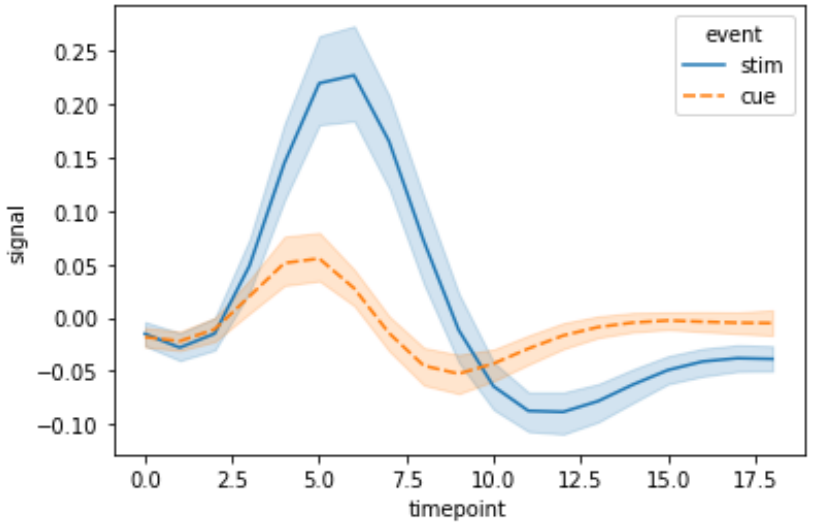
```
[94]: sns.relplot(data=fmri,x="timepoint", y="signal",hue="event", style="event",kind="line")
```

```
[94]: <seaborn.axisgrid.FacetGrid at 0x1dc93bc05c8>
```




```
sns.lineplot(data=fmri,x="timepoint",hue="event", style="event",y="signal")
```

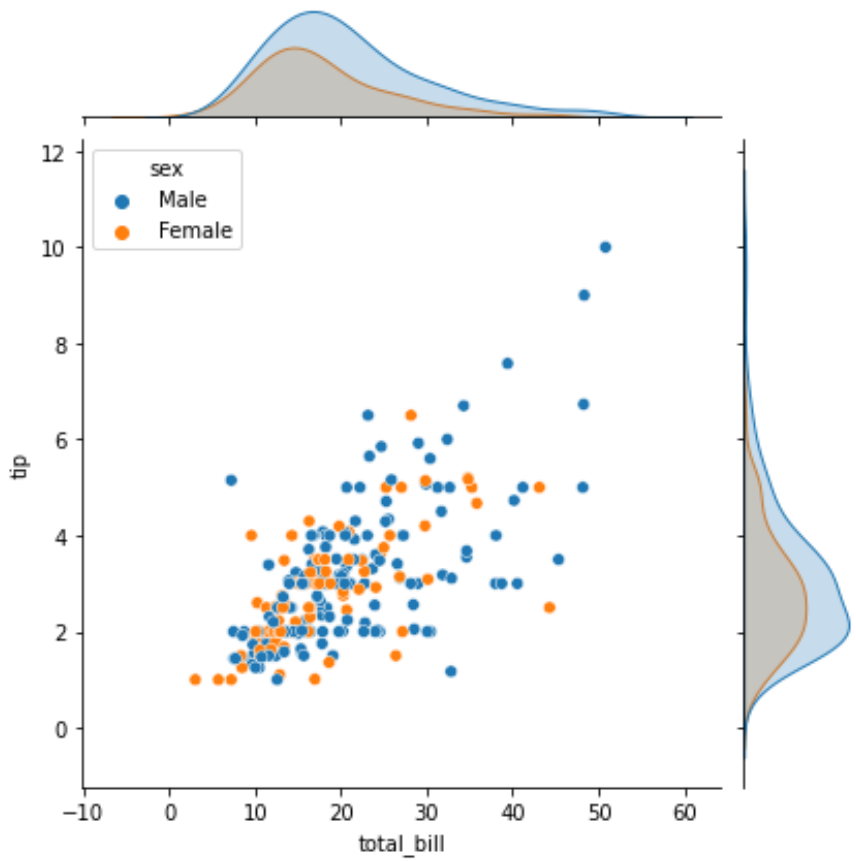
```
<AxesSubplot:xlabel='timepoint', ylabel='signal'>
```



jointplot&pairplot

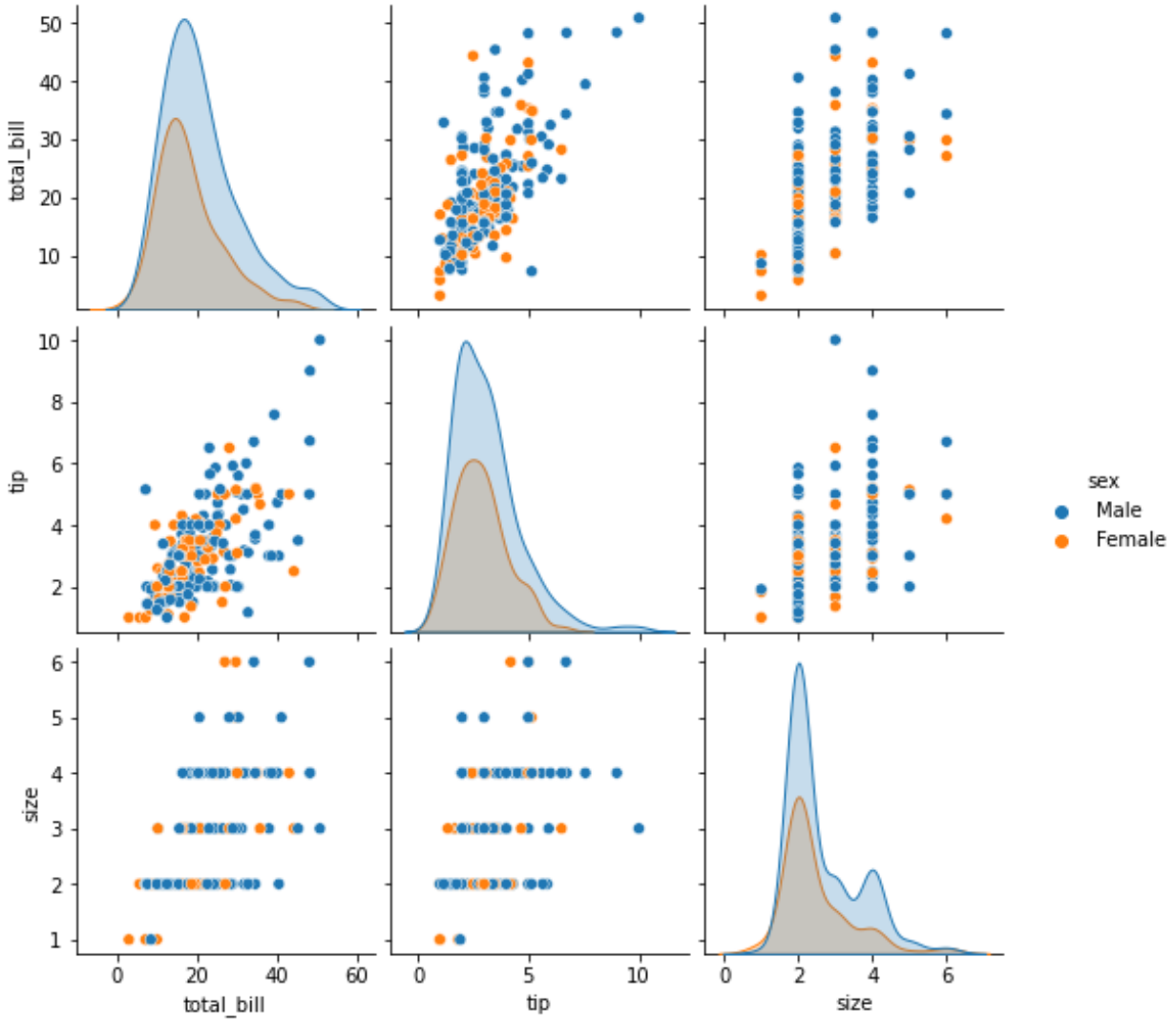
```
sns.jointplot(data=tips, x="total_bill", y="tip",hue='sex')
```

```
<seaborn.axisgrid.JointGrid at 0x1dc945061c8>
```




```
sns.pairplot(data=tips,hue='sex')
```

```
<seaborn.axisgrid.PairGrid at 0x1dc95c4f348>
```

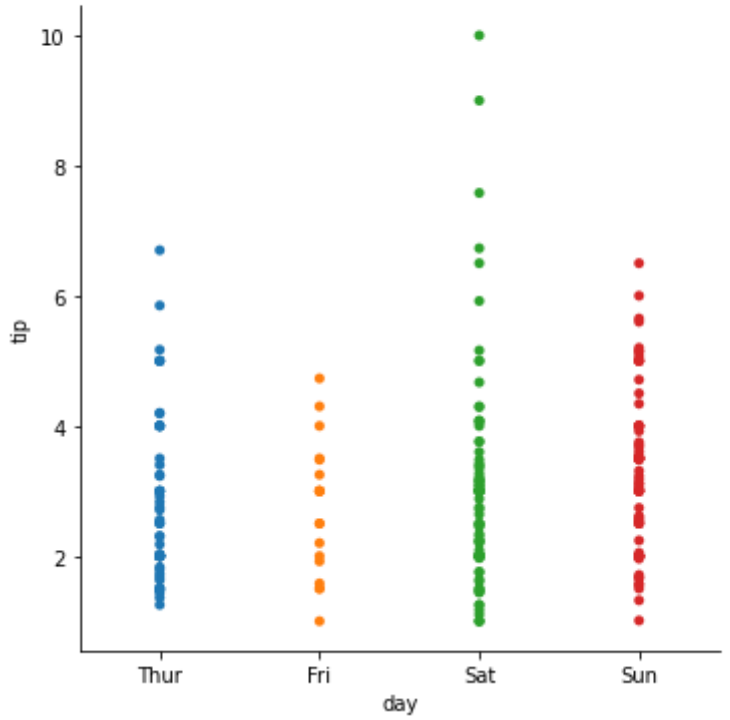


catplot-scatter

```
sns.stripplot(data=tips,x="day",jitter=False,y="tip")
```

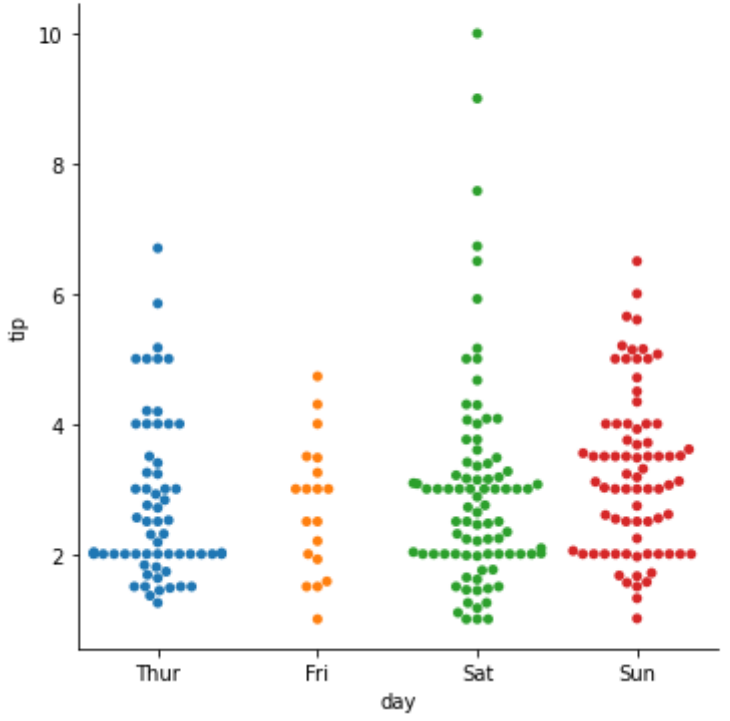
```
sns.catplot(data=tips,x="day",jitter=False,y="tip")
```

```
<seaborn.axisgrid.FacetGrid at 0x1dc96423bc8>
```




```
sns.catplot(data=tips,x="day", y="tip", kind="swarm")
```

```
<seaborn.axisgrid.FacetGrid at 0x1dc964fea88>
```

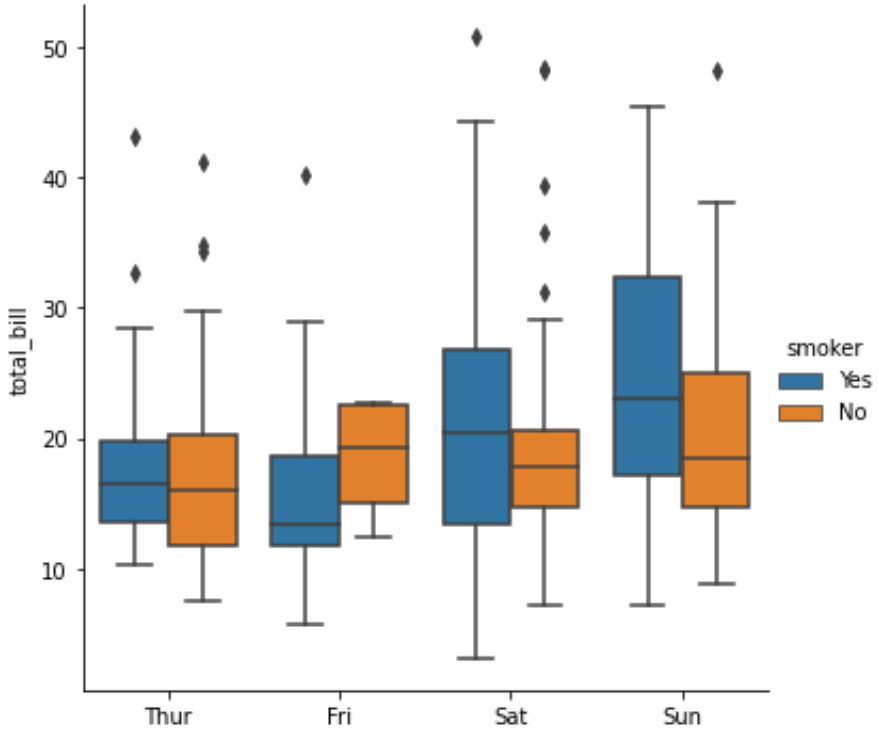


catplot-distribute

根据分位数分箱，box四分位数，boxen更多的分位数

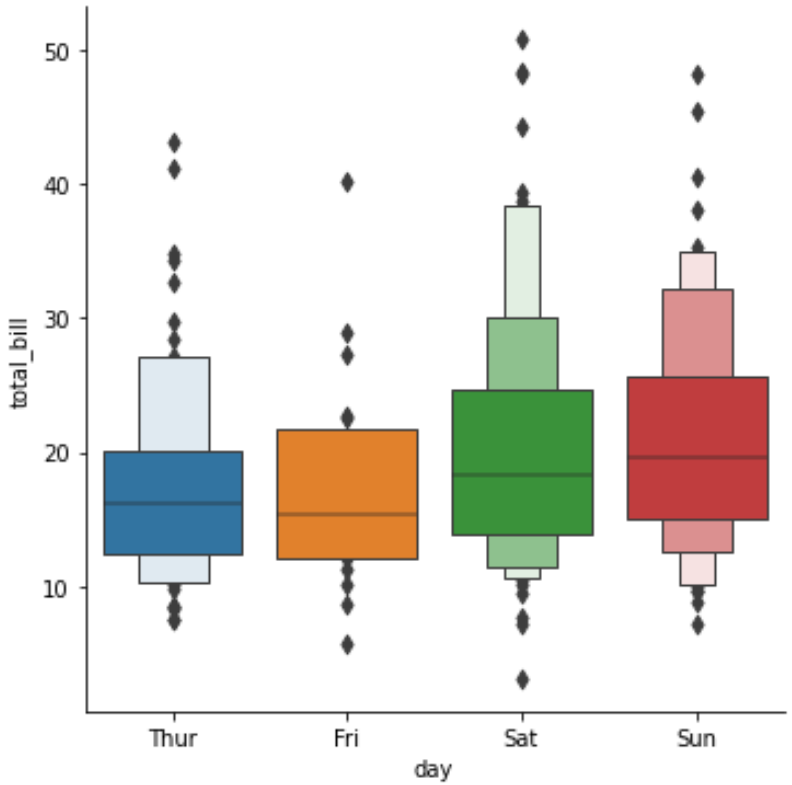
```
sns.catplot(x="day", y="total_bill", kind="box", data=tips,hue="smoker")
```

```
<seaborn.axisgrid.FacetGrid at 0x1dc977fd108>
```




```
sns.catplot(x="day", y="total_bill", kind="boxen", data=tips)
```

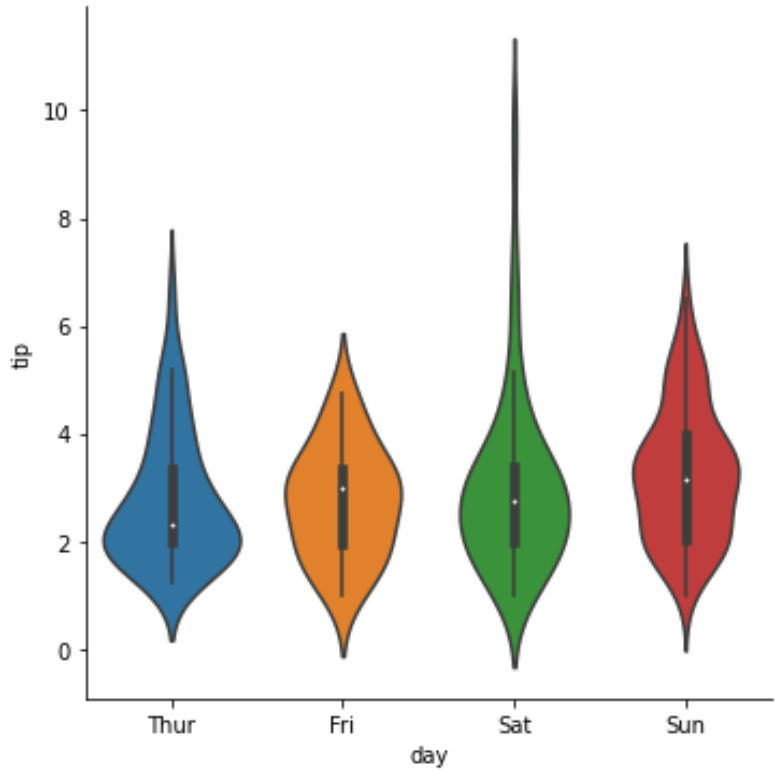
```
<seaborn.axisgrid.FacetGrid at 0x1dc967963c8>
```



catplot-distribute

```
sns.catplot(x="day", y="tip", kind="violin", data=tips)
```

```
<seaborn.axisgrid.FacetGrid at 0x1dc97979248>
```

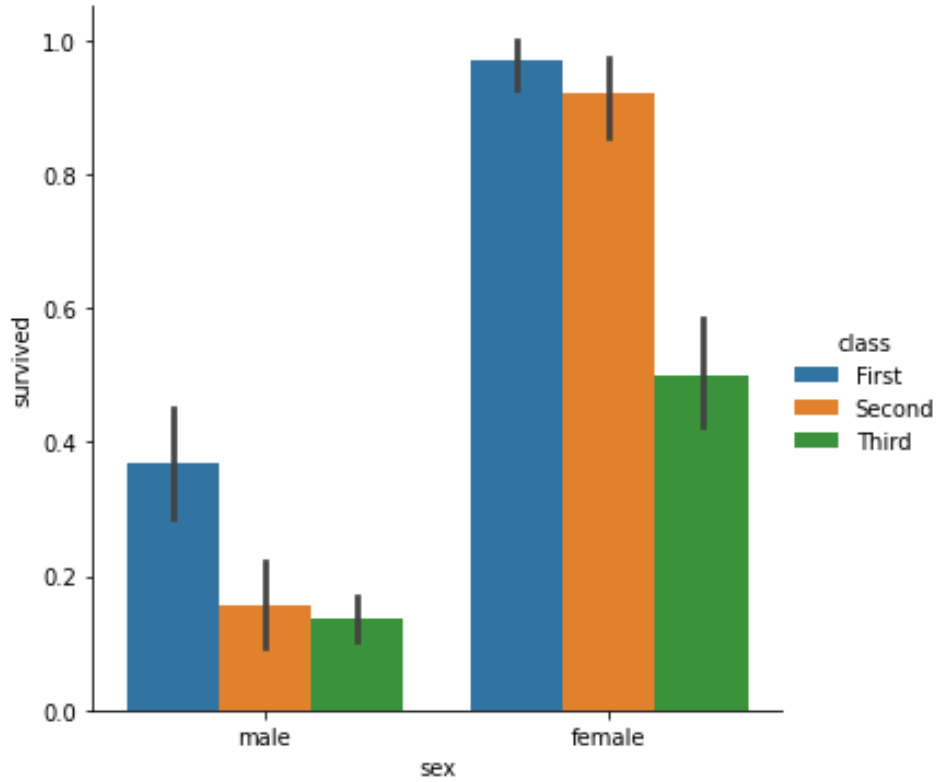


概率密度图和box的结合，因此可使用和kde类似的平滑参数

catplot -statistic

```
sns.catplot(x="sex", y="survived", hue="class", kind="bar", data=titanic)
```

```
<seaborn.axisgrid.FacetGrid at 0x1dc97b73bc8>
```



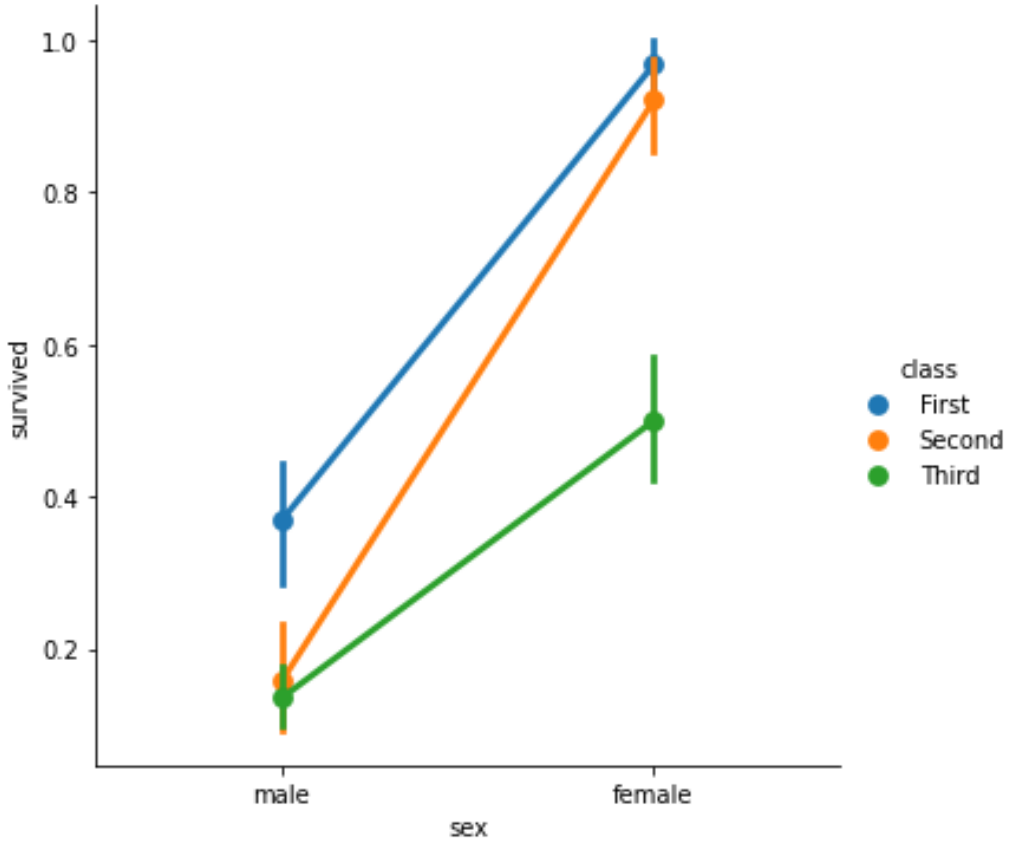

```
titanic = sns.load_dataset("titanic")
titanic
```

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class
0	0	3	male	22.0	1	0	7.2500	S	Third
1	1	1	female	38.0	1	0	71.2833	C	First
2	1	3	female	26.0	0	0	7.9250	S	Third
3	1	1	female	35.0	1	0	53.1000	S	First
4	0	3	male	35.0	0	0	8.0500	S	Third

catplot -statistic

```
sns.catplot(x="sex", y="survived", hue="class", kind="point", data=titanic)
```

```
<seaborn.axisgrid.FacetGrid at 0x1dc97bd1408>
```



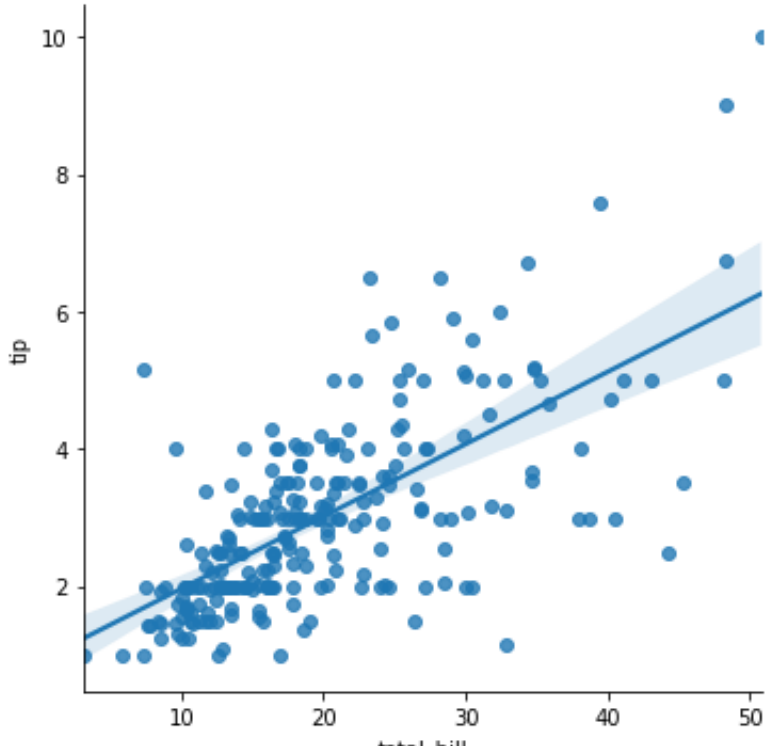
通过线条的起伏反映变化

回归分析

```
sns.regplot(x="total_bill", y="tip", data=tips)
```

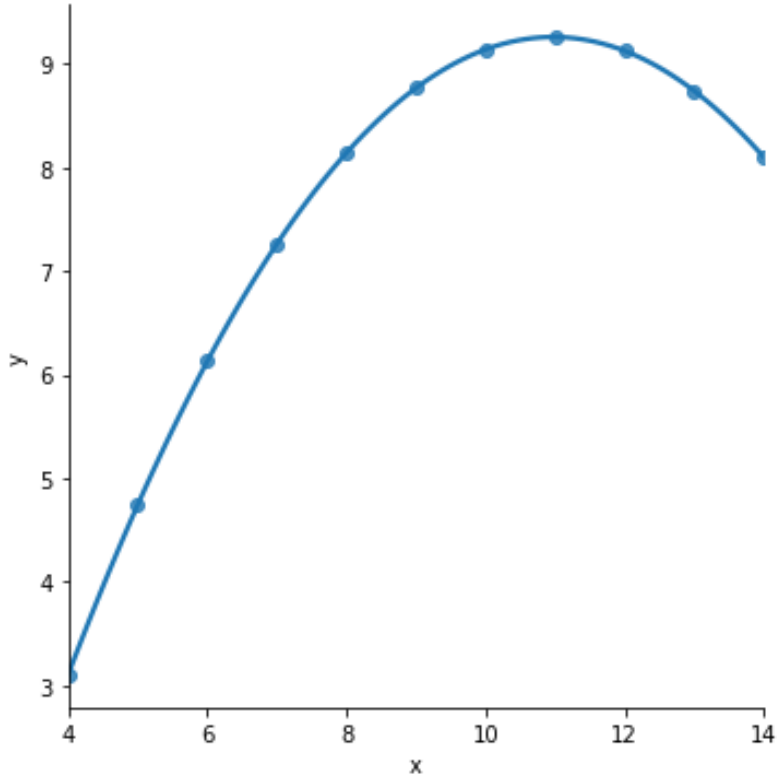
```
sns.lmplot(x="total_bill", y="tip", data=tips)
```

```
<seaborn.axisgrid.FacetGrid at 0x1dc97d7f6c8>
```




```
anscombe = sns.load_dataset("anscombe")
sns.lmplot(x="x", y="y", data=anscombe.query("dataset == 'II'"),order=2)
```

```
<seaborn.axisgrid.FacetGrid at 0x1dc9a300108>
```

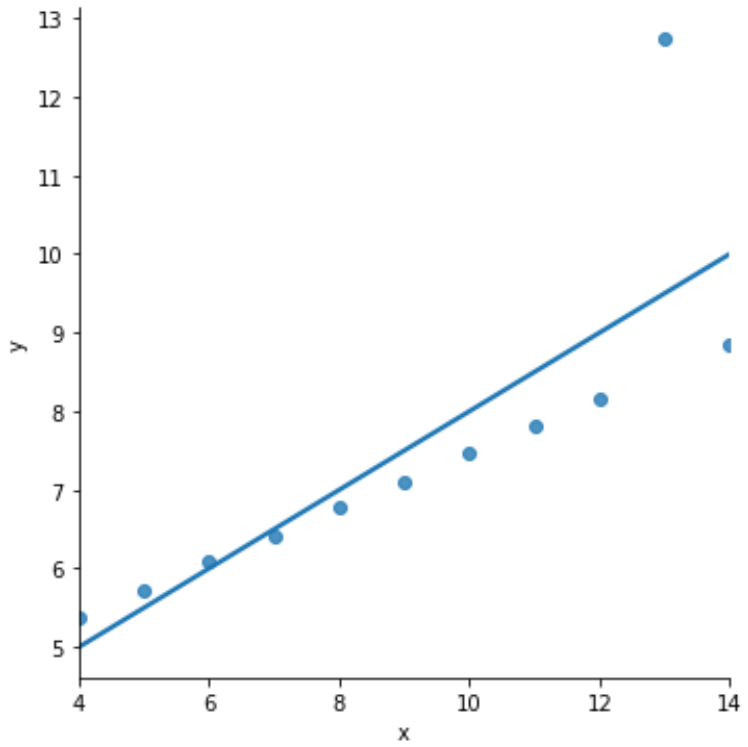


回归分析

通过robust选项忽略特异点

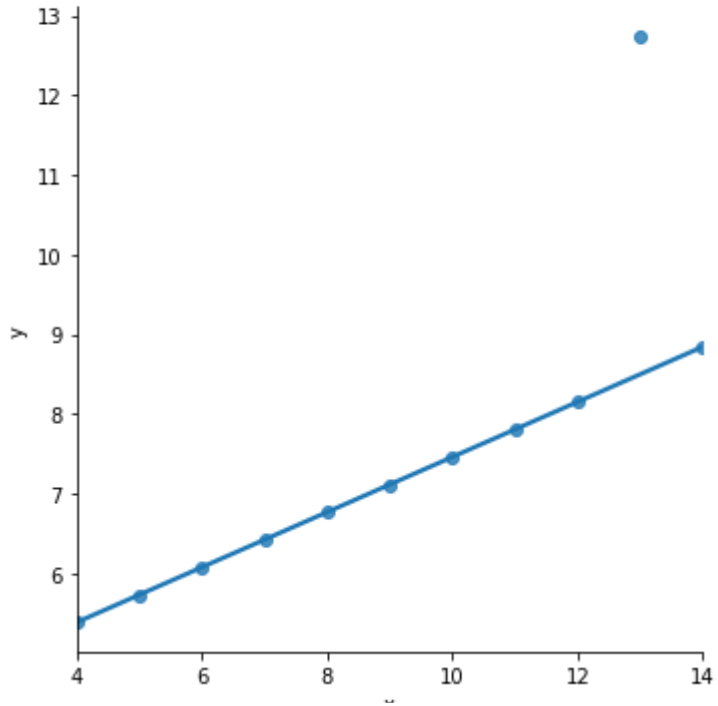
```
sns.lmplot(x="x", y="y", data=anscombe.query("dataset == 'III'"),ci=None)
```

```
<seaborn.axisgrid.FacetGrid at 0x1dc9a3d0a88>
```




```
sns.lmplot(x="x", y="y", data=anscombe.query("dataset == 'III'"),ci=None,robust=True)
```

```
<seaborn.axisgrid.FacetGrid at 0x1dc9a3d1288>
```

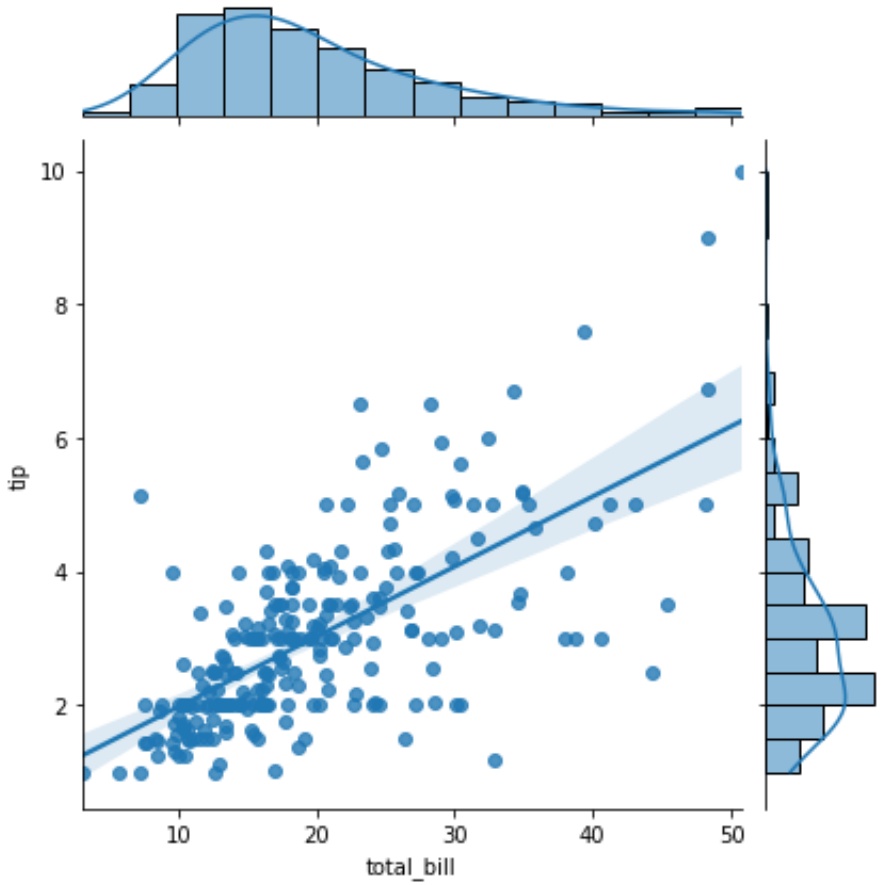


回归分析

通过kind参数可以在其他图形级绘图函数中进行回归分析

```
sns.jointplot(x="total_bill", y="tip", data=tips, kind="reg")
```

```
<seaborn.axisgrid.JointGrid at 0x1dc9a6b2ac8>
```



总结

- 画布级函数：
  - displot,relplot,catplot
  - jointplot,pairplot,lmplot

- 参数：
  - col,row: 分类子图（画布级）
  - kind: 图形类别 （画布级）
  - hue,style: 颜色，样式区分
  - ...
  - 详细信息：http://seaborn.pydata.org/api.html

练一练

- 绘制iris数据集的联合分布和边缘分布，联合分布使用回归分析；

  - 提示：提交你的notebook

  1.加载数据集 load_dataset,
  2.多变量分布关系pair_plot;

THANKS