1 第三章：绘图

数据小鱼Rexa
CSDN：https://blog.csdn.net/qq_38395376?spm=1011.2124.3001.5343
Blibli：https://space.bilibili.com/283181288
Github：https://github.com/Rexa-Yu

1.1 读取数据

In [38]:

```
import seaborn as sns
import pandas as pd
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')
%matplotlib inline
anscombe=pd.read_csv(r'E:\jupyter notebook storage\Practice in Pandas\seaborn-data-m
anscombe.head()
```

Out[38]:

	dataset	x	y
0	I	10.0	8.04
1	I	8.0	6.95
2	I	13.0	7.58
3	I	9.0	8.81
4	I	11.0	8.33

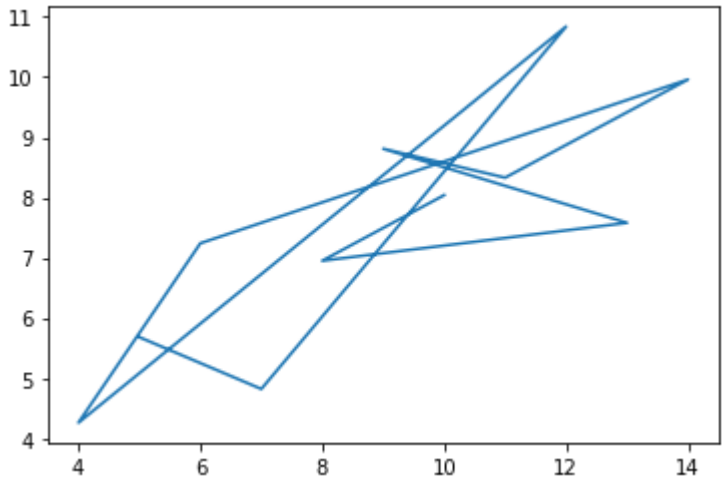
1.2 基础画图

In [39]:

```
dataset1=anscombe[anscombe['dataset']=="I"]
plt.plot(dataset1['x'],dataset1["y"])
```

Out[39]:

```
[<matplotlib.lines.Line2D at 0x1e84548f220>]
```

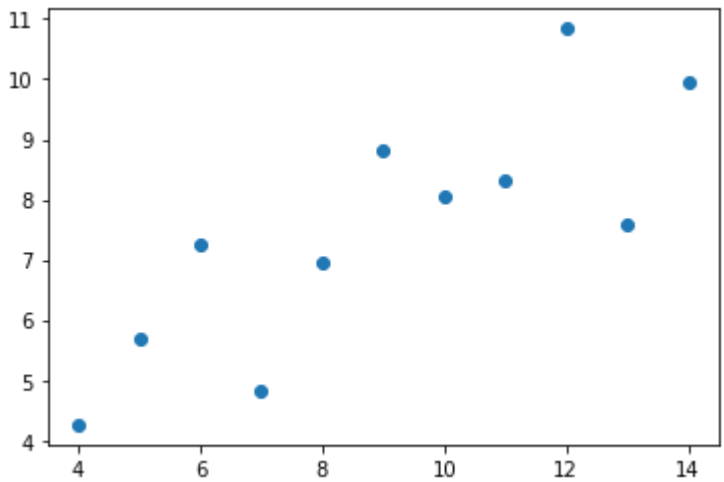


In [40]:

```
# 如果需要点图，则需要传参"o"
plt.plot(dataset1["x"],dataset1["y"],"o")
```

Out[40]:

```
[<matplotlib.lines.Line2D at 0x1e845460610>]
```



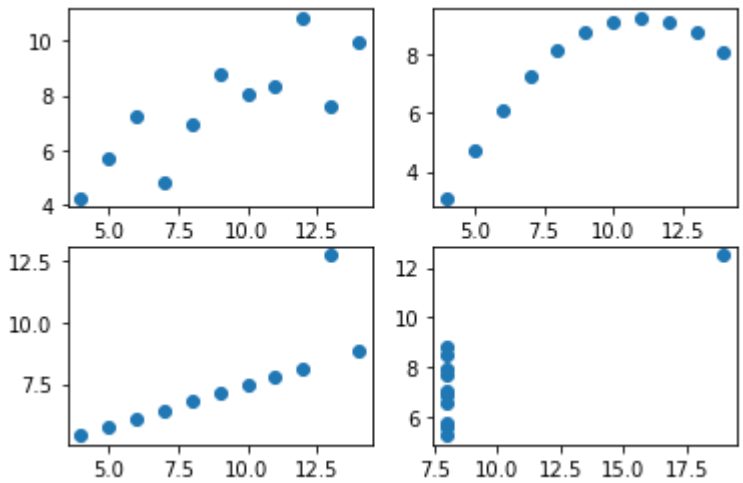
1.3 画布画图

In [41]:

```
# 创建画布
fig=plt.figure()
axe1=fig.add_subplot(2,2,1)
axe2=fig.add_subplot(2,2,2)
axe3=fig.add_subplot(2,2,3)
axe4=fig.add_subplot(2,2,4)
# 定义数据
dataset2=anscombe[anscombe["dataset"]=="II"]
dataset3=anscombe[anscombe["dataset"]=="III"]
dataset4=anscombe[anscombe["dataset"]=="IV"]
# 将4个子画布，分别代入数据
axe1.plot(dataset1["x"],dataset1["y"],"o")
axe2.plot(dataset2["x"],dataset2["y"],"o")
axe3.plot(dataset3["x"],dataset3["y"],"o")
axe4.plot(dataset4["x"],dataset4["y"],"o")
```

Out[41]:

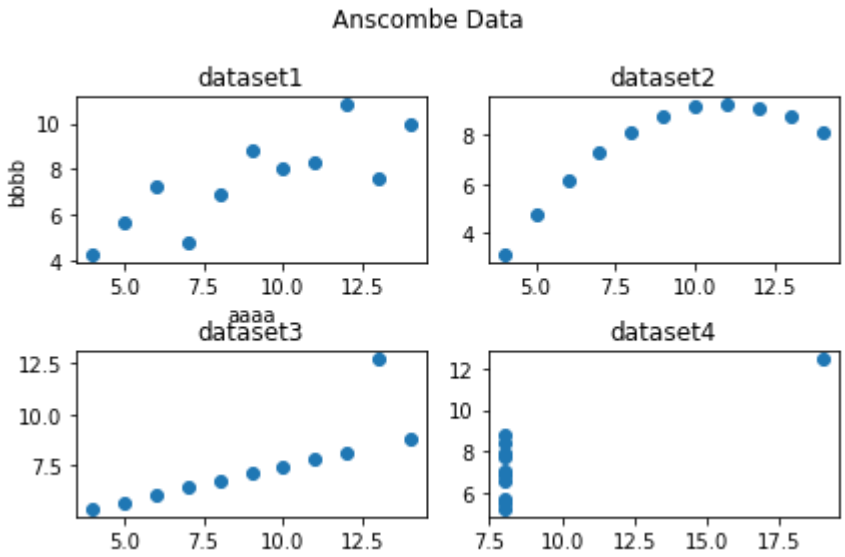
```
[<matplotlib.lines.Line2D at 0x1e8456f7c40>]
```



In [42]:

```
# 添加标题
axe1.set_title("dataset1")
axe2.set_title("dataset2")
axe3.set_title("dataset3")
axe4.set_title("dataset4")
# 为画布添加大标题
fig.suptitle("Anscombe Data")
# 使用紧凑型布局
fig.tight_layout()
# 设置横坐标
axe1.set_xlabel("aaaa")
axe1.set_ylabel("bbbb")
fig
```

Out[42]:



1.4 制作统计图（柱形图）

In [43]:

```
# 加载数据集
tips=pd.read_csv("E:\jupyter notebook storage\Practice in Pandas\seaborn-data-master
tips
```

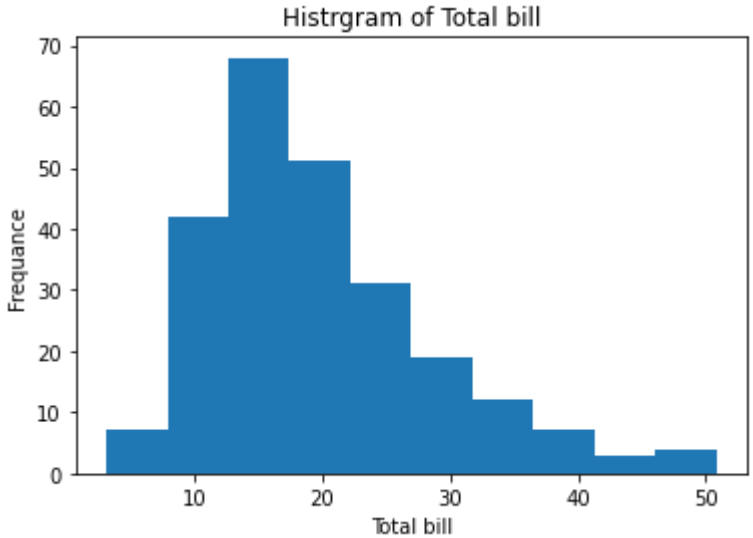
Out[43]:

	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

244 rows × 7 columns

In [44]:

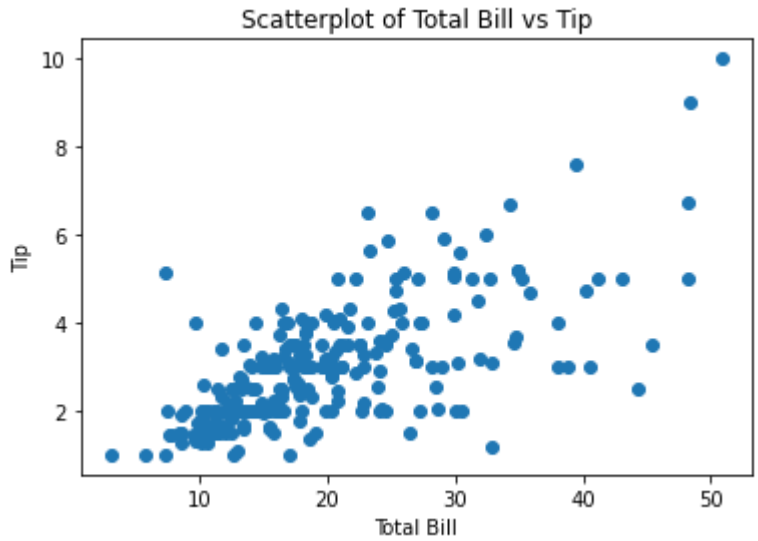
```
fig=plt.figure()
axes1=fig.add_subplot(1,1,1)
axes1.hist(tips["total_bill"],bins=10)
axes1.set_title("Histrgram of Total bill")
axes1.set_xlabel("Total bill")
axes1.set_ylabel("Frequance")
fig.show()
```



1.5 散点图（双变量）

In [45]:

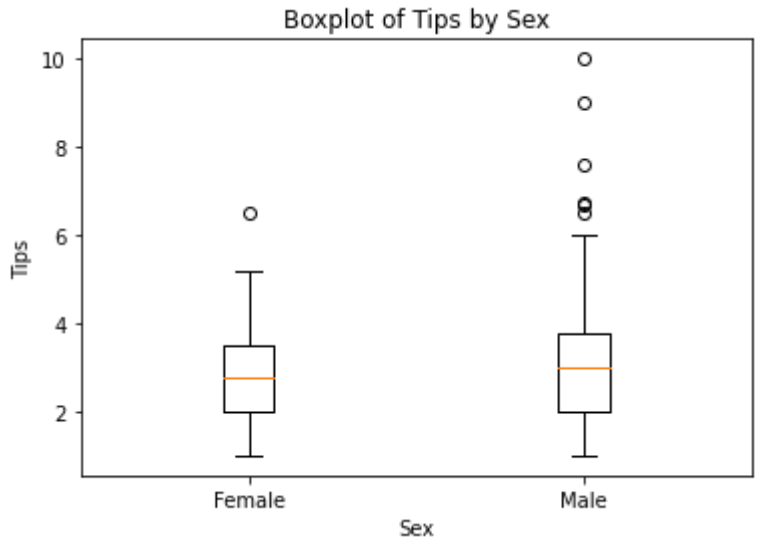
```
scatter_plot=plt.figure()
axes1=scatter_plot.add_subplot(1,1,1)
axes1.scatter(tips["total_bill"],tips["tip"])
axes1.set_title("Scatterplot of Total Bill vs Tip")
axes1.set_xlabel("Total Bill")
axes1.set_ylabel("Tip")
scatter_plot.show()
```



1.6 箱线图

In [46]:

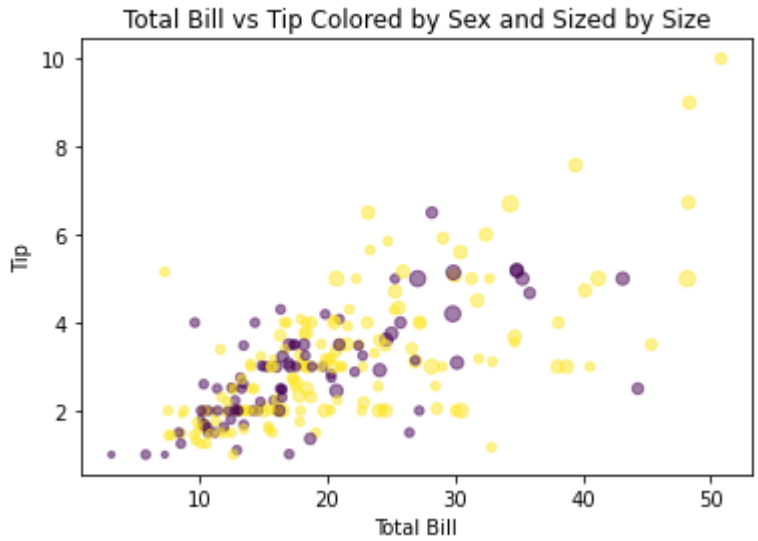
```
boxplot=plt.figure()
axes1=boxplot.add_subplot(1,1,1)
axes1.boxplot([tips[tips["sex"]=="Female"]["tip"],tips[tips["sex"]=="Male"]["tip"]],
axes1.set_xlabel("Sex")
axes1.set_ylabel("Tips")
axes1.set_title("Boxplot of Tips by Sex")
boxplot.show()
```



1.7 多变量数据

In [47]:

```
# 基于性别创建一个带颜色函数
def recode_sex(sex):
    if sex=="Female":
        return 0
    else:
        return 1
tips["sex_color"]=tips["sex"].apply(recode_sex)
scatter_plot=plt.figure()
axes1=scatter_plot.add_subplot(1,1,1)
axes1.scatter(x=tips["total_bill"],y=tips["tip"],s=tips["size"]*10,c=tips["sex_color
axes1.set_title("Total Bill vs Tip Colored by Sex and Sized by Size")
axes1.set_xlabel("Total Bill")
axes1.set_ylabel("Tip")
scatter_plot.show()
```

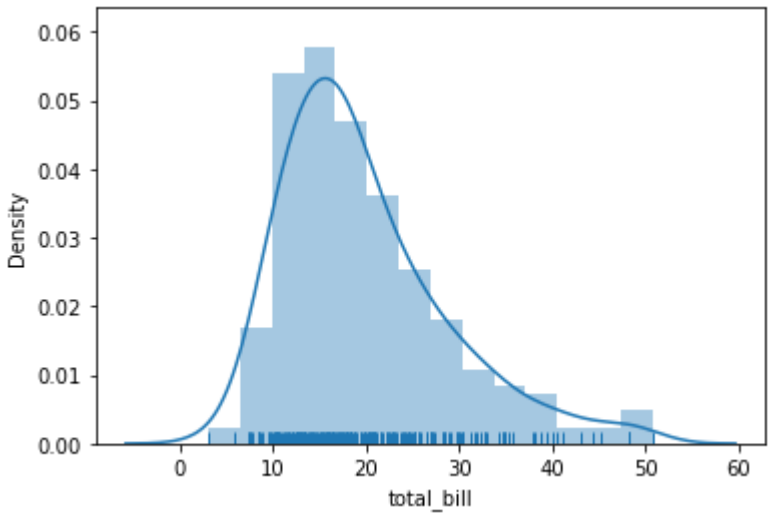


1.8 seaborn 画图

1.8.1 直方图

In [48]:

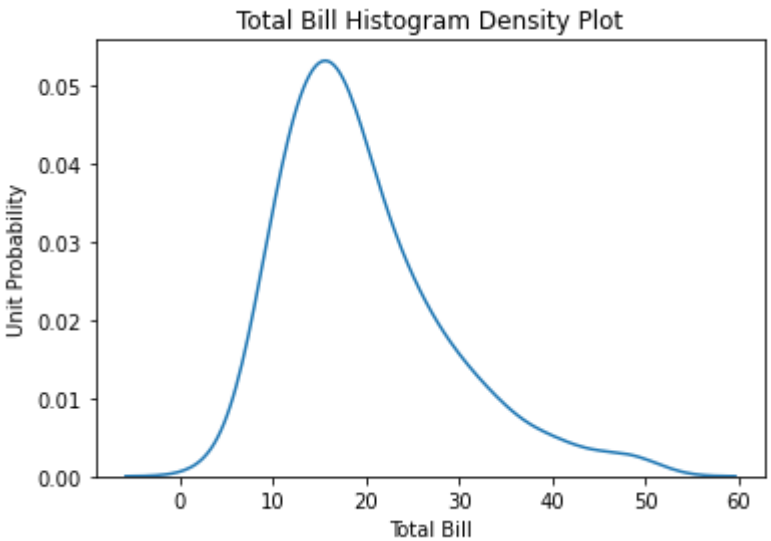
```
hist,ax=plt.subplots()
ax=sns.distplot(tips["total_bill"],kde=True,rug=True)
plt.show()
```



1.8.2 密度图

In [49]:

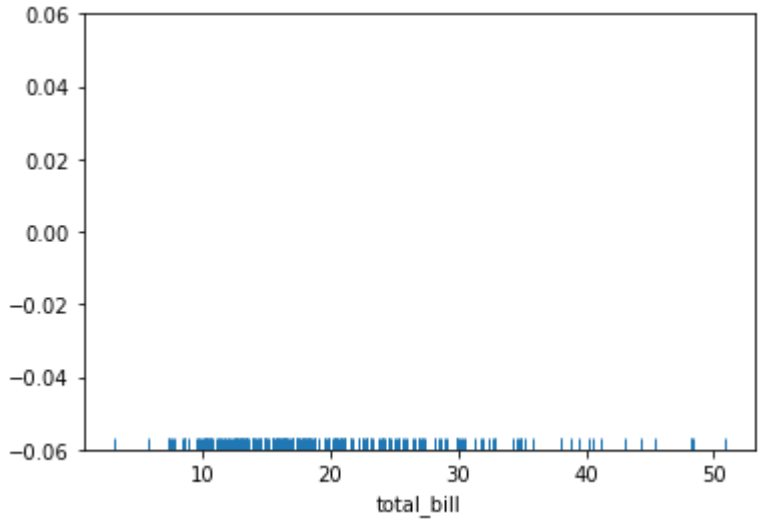
```
ax=plt.subplots()
ax=sns.kdeplot(tips["total_bill"])
ax.set_title("Total Bill Histogram Density Plot")
ax.set_xlabel("Total Bill")
ax.set_ylabel("Unit Probability")
plt.show()
```



1.8.3 频数图

In [50]:

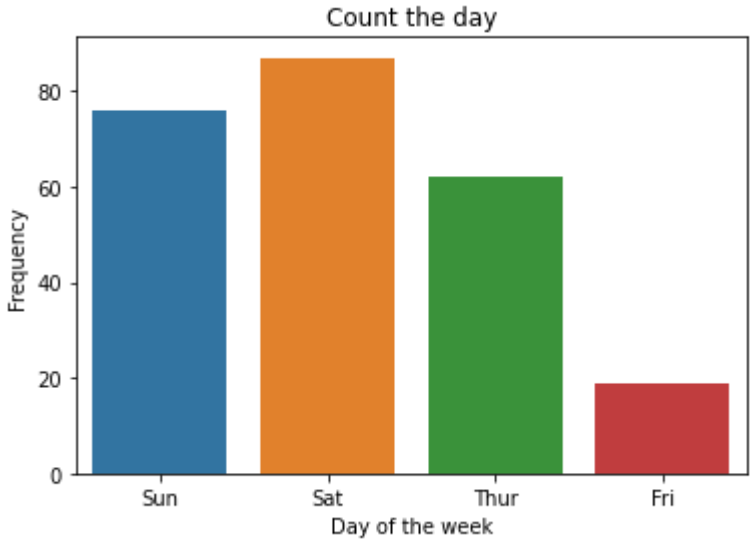
```
ax=plt.subplots()
ax=sns.rugplot(tips["total_bill"])
plt.show()
```



1.8.4 计数图

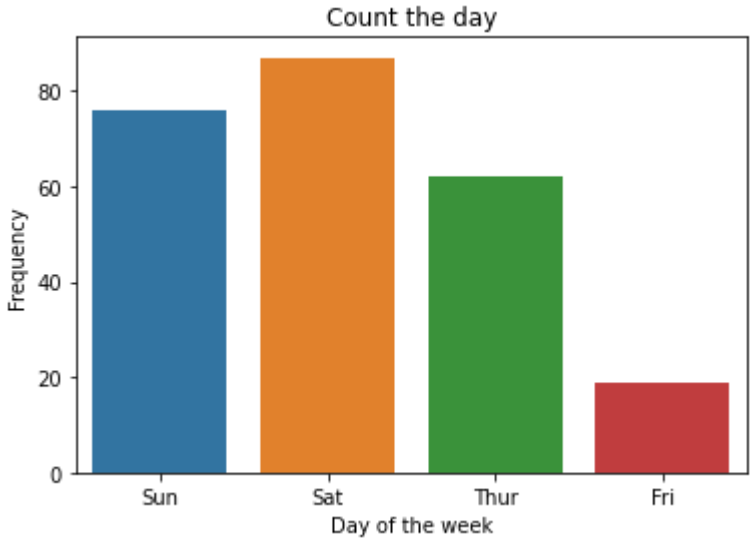
In [51]:

```
ax=plt.subplots()
ax=sns.countplot("day",data=tips)
ax.set_title("Count the day")
ax.set_xlabel("Day of the week")
ax.set_ylabel("Frequency")
plt.show()
```



In [52]:

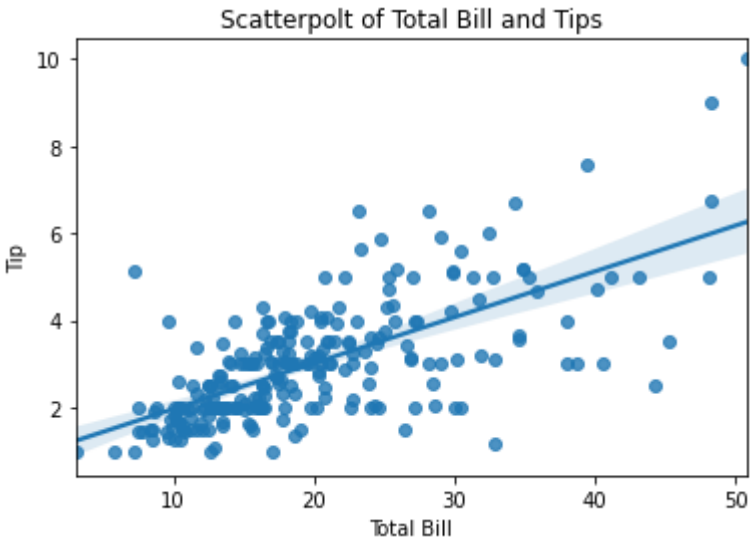
```
ax=plt.subplots()
ax=sns.countplot(tips["day"])
ax.set_title("Count the day")
ax.set_xlabel("Day of the week")
ax.set_ylabel("Frequency")
plt.show()
```



1.8.5 散点图

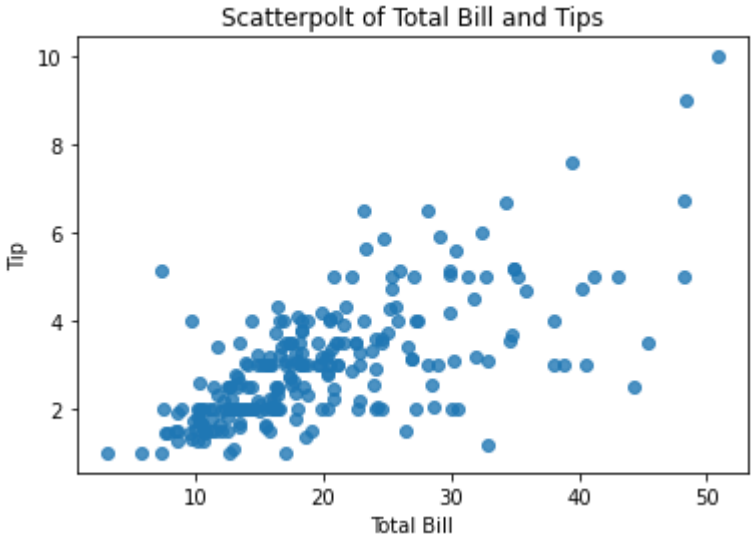
In [53]:

```
ax=plt.subplots()
ax=sns.regplot(x="total_bill",y="tip",data=tips)
ax.set_title("Scatterpolt of Total Bill and Tips")
ax.set_xlabel("Total Bill")
ax.set_ylabel("Tip")
plt.show()
```



In [54]:

```
# fit_reg=False，则不会有拟合线。
ax=plt.subplots()
ax=sns.regplot(x="total_bill",y="tip",data=tips,fit_reg=False)
ax.set_title("Scatterpolt of Total Bill and Tips")
ax.set_xlabel("Total Bill")
ax.set_ylabel("Tip")
plt.show()
```



In [55]:

```
# 查看数据集
tips
```

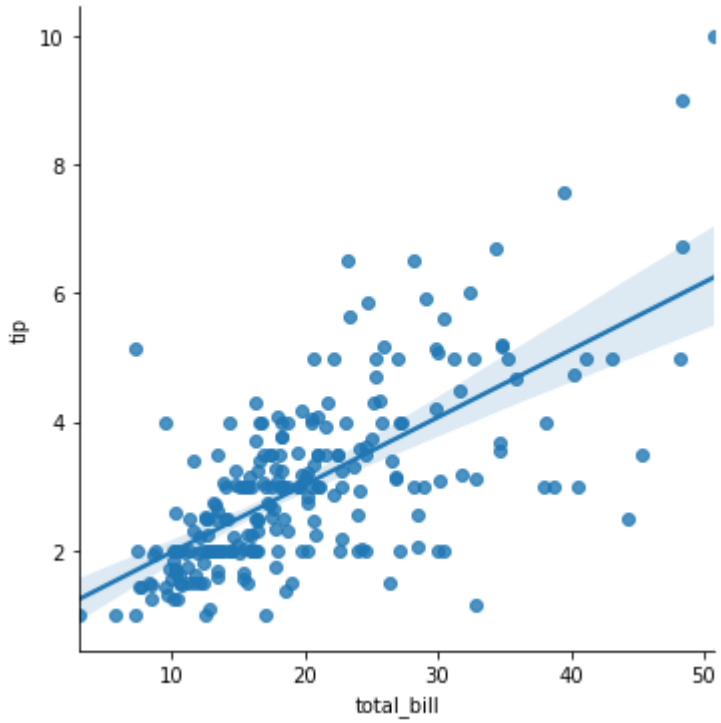
Out[55]:

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

244 rows × 8 columns

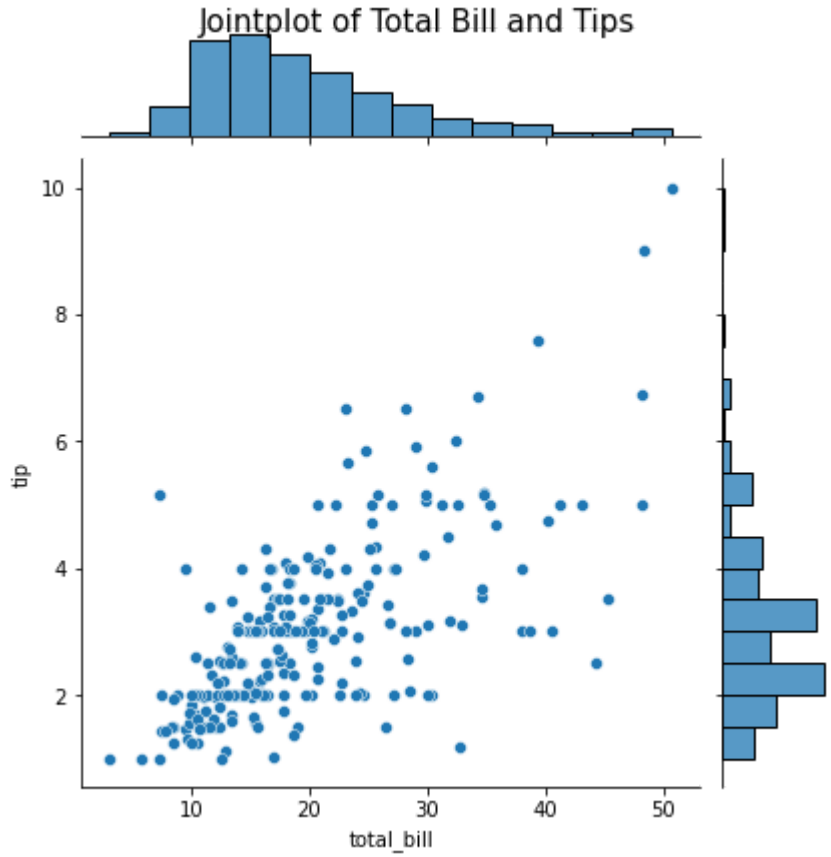
In [56]:

```
# 同样的，可以使用implot函数实施散点图，不同点在于，regplot是在轴域上画图，implot实在画布上，也就
fig=sns.lmplot(x="total_bill",y="tip",data=tips)
plt.show()
```



In [57]:

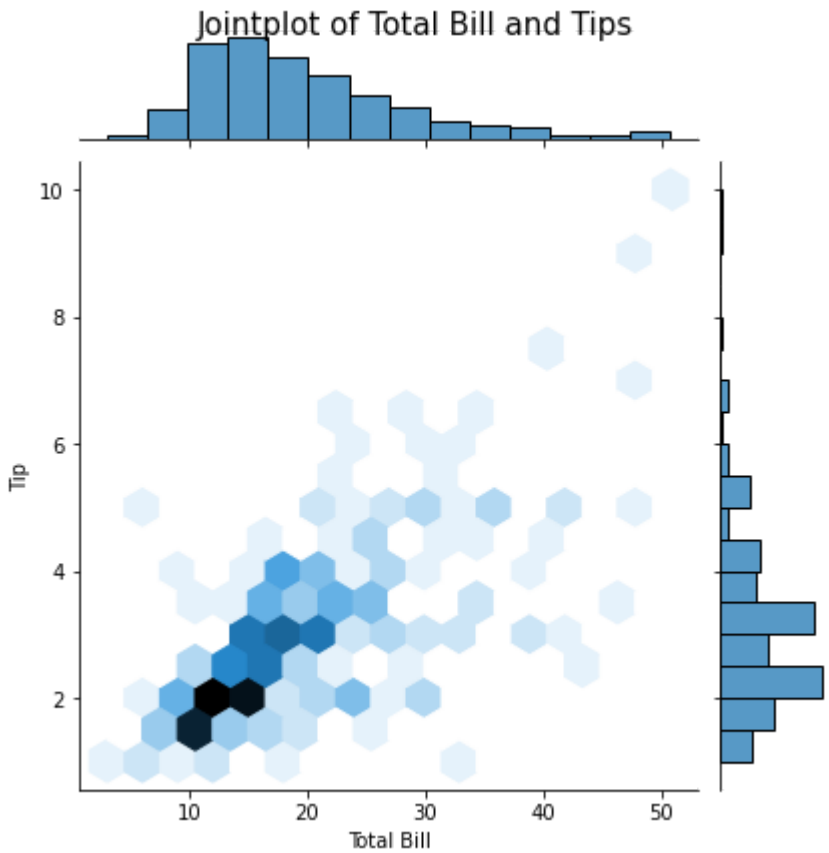
```
#  如果需要看单个变量的散点图，即直方图，可以用jointplot来操作
joint=sns.jointplot(x="total_bill",y="tip",data=tips)
# 特殊地，它的标题不以set_titel函数添加，用fig.suptitle()画布函数标题来添加
joint.fig.suptitle("Jointplot of Total Bill and Tips",fontsize=15,y=1)
plt.show()
```



1.8.6 蜂巢图

In [58]:

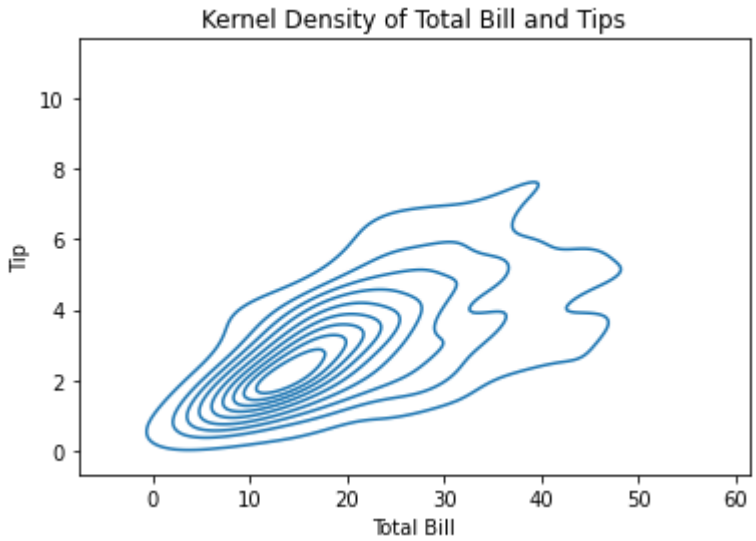
```
# 蜂巢图可以直观看到一类数据的频次
hexbin=sns.jointplot(x="total_bill",y="tip",data=tips,kind="hex")
# 设置x，y的标签
hexbin.set_axis_labels(xlabel="Total Bill",ylabel="Tip")
# 特殊地，它的标题不以set_titel函数添加，用fig.suptitle()画布函数标题来添加
hexbin.fig.suptitle("Jointplot of Total Bill and Tips",fontsize=15,y=1)
plt.show()
```



1.8.7 2D密度图

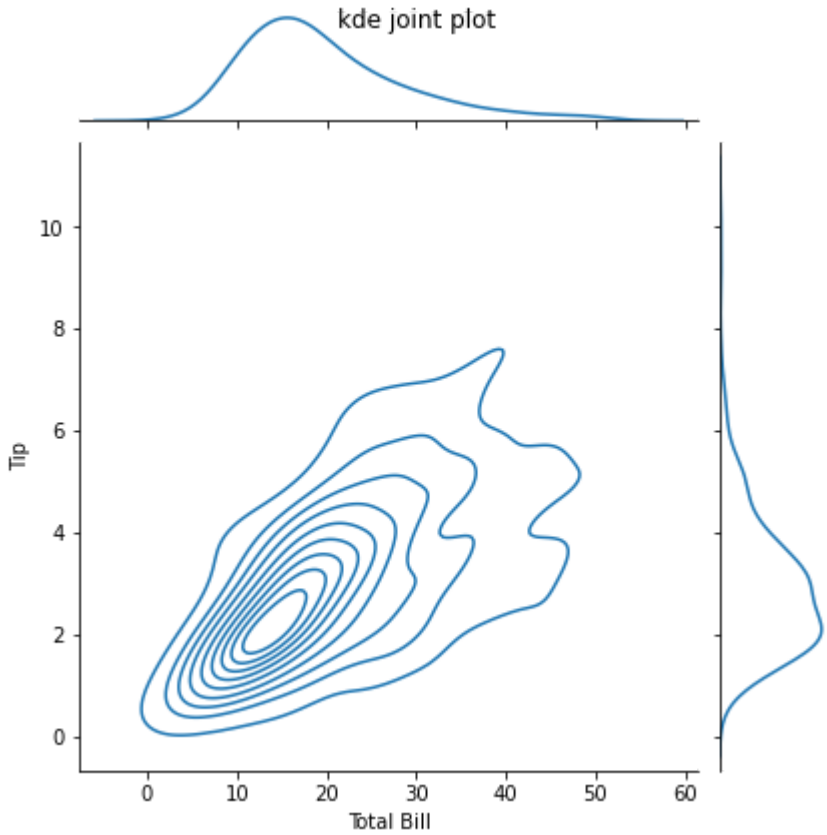
In [59]:

```
ax=plt.subplots()
ax=sns.kdeplot(data=tips["total_bill"],data2=tips["tip"],shad=True)
ax.set_xlabel("Total Bill")
ax.set_ylabel("Tip")
ax.set_title("Kernel Density of Total Bill and Tips")
plt.show()
```



In [60]:

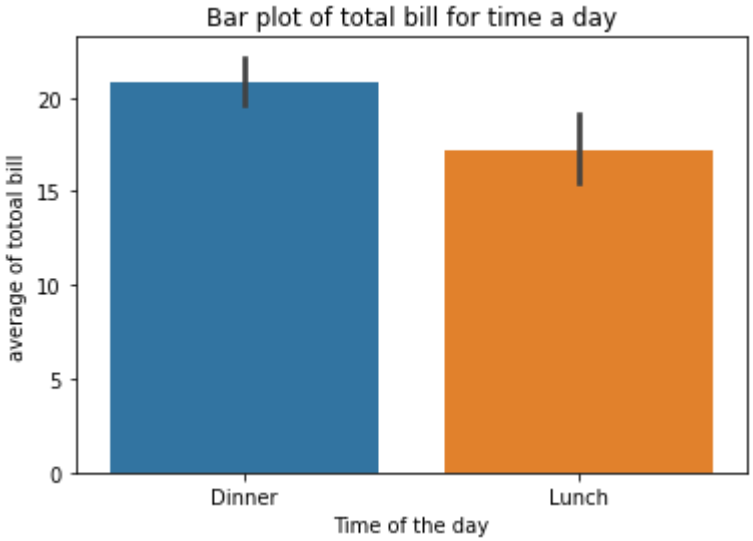
```
#  同样的效果，jointplot也可以实现
ked_joint=sns.jointplot(x="total_bill",y="tip",data=tips,kind="kde")
ked_joint.fig.suptitle("kde joint plot")
ked_joint.set_axis_labels(xlabel="Total Bill",ylabel="Tip")
plt.show()
```



1.8.8 条形图

In [61]:

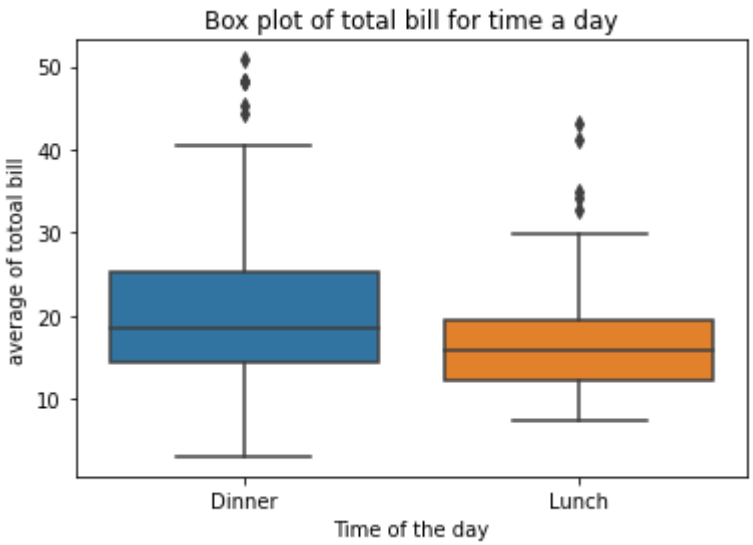
```
# 条形图画图时自动计算平均值，因此在观察数据均值可以使用条形图观察均值以及数据分布量
ax=plt.subplots()
ax=sns.barplot(x="time",y="total_bill",data=tips)
ax.set_xlabel("Time of the day")
ax.set_ylabel("average of totoal bill")
ax.set_title("Bar plot of total bill for time a day")
plt.show()
```



1.8.9 箱线图

In [62]:

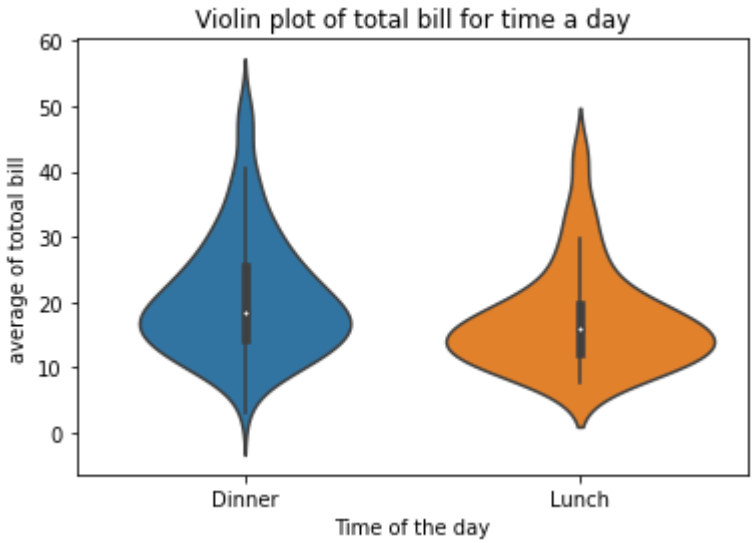
```
ax=plt.subplots()
ax=sns.boxplot(x="time",y="total_bill",data=tips)
ax.set_xlabel("Time of the day")
ax.set_ylabel("average of totoal bill")
ax.set_title("Box plot of total bill for time a day")
plt.show()
```



1.8.10 小提琴图

In [63]:

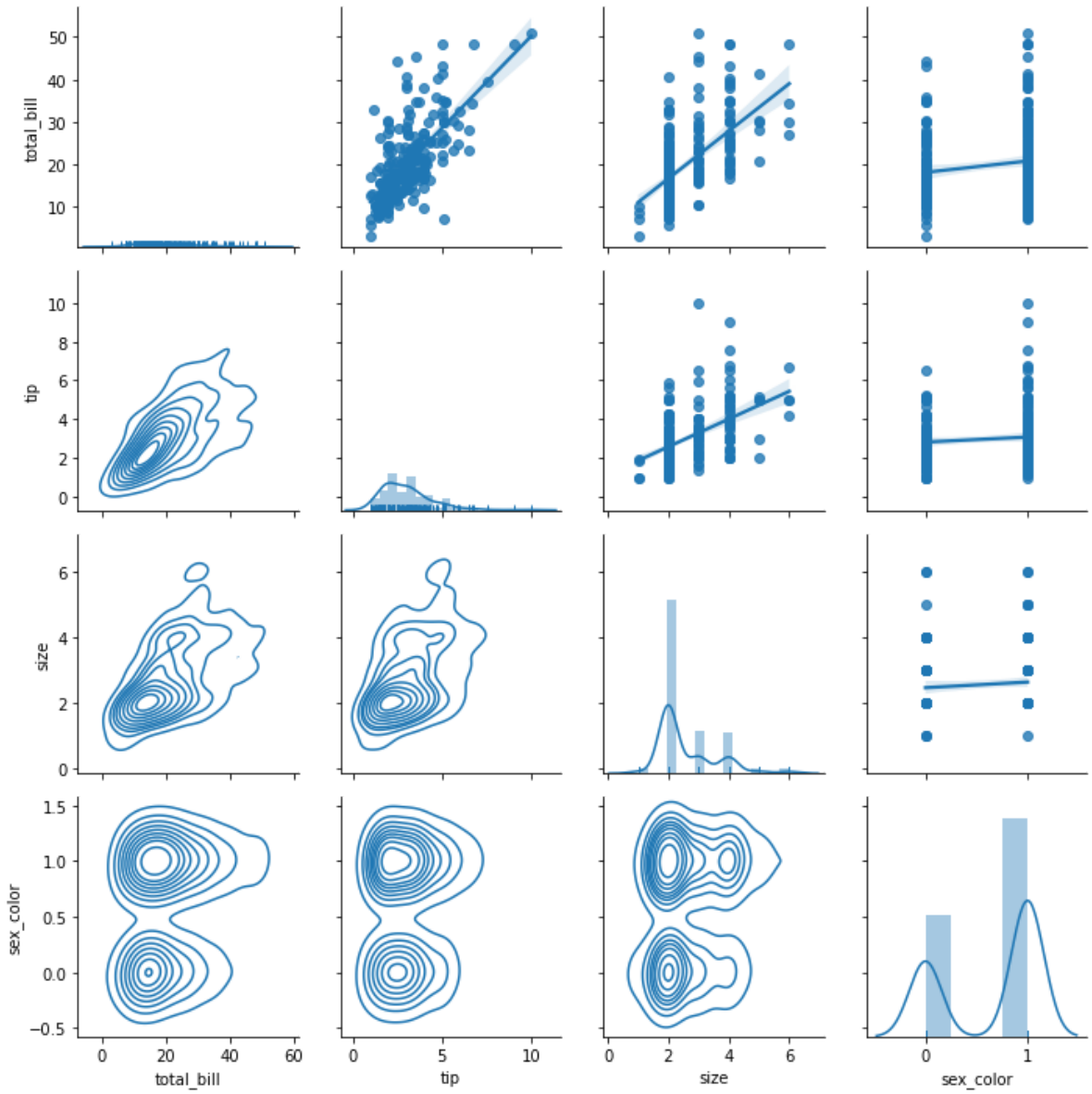
```
ax=plt.subplots()
ax=sns.violinplot(x="time",y="total_bill",data=tips)
ax.set_xlabel("Time of the day")
ax.set_ylabel("average of totoal bill")
ax.set_title("Violin plot of total bill for time a day")
plt.show()
```



1.8.11 成对关系图

In [64]:

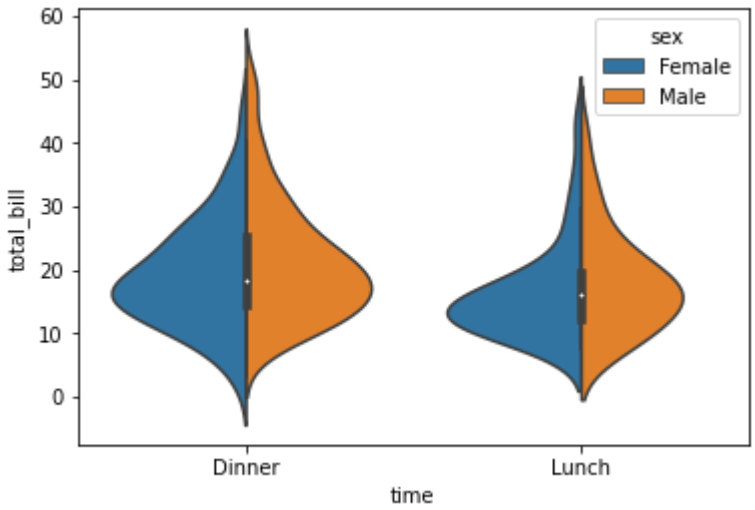
```
pair_grid=sns.PairGrid(tips)
# 先画中间的，即map_diag函数，然后在画上下图，这样才能正常显示
pair_grid=pair_grid.map_diag(sns.distplot,kde="True",rug="True")
pair_grid=pair_grid.map_upper(sns.regplot)
pair_grid=pair_grid.map_lower(sns.kdeplot)

plt.show()
```



1.8.12 多变量数据

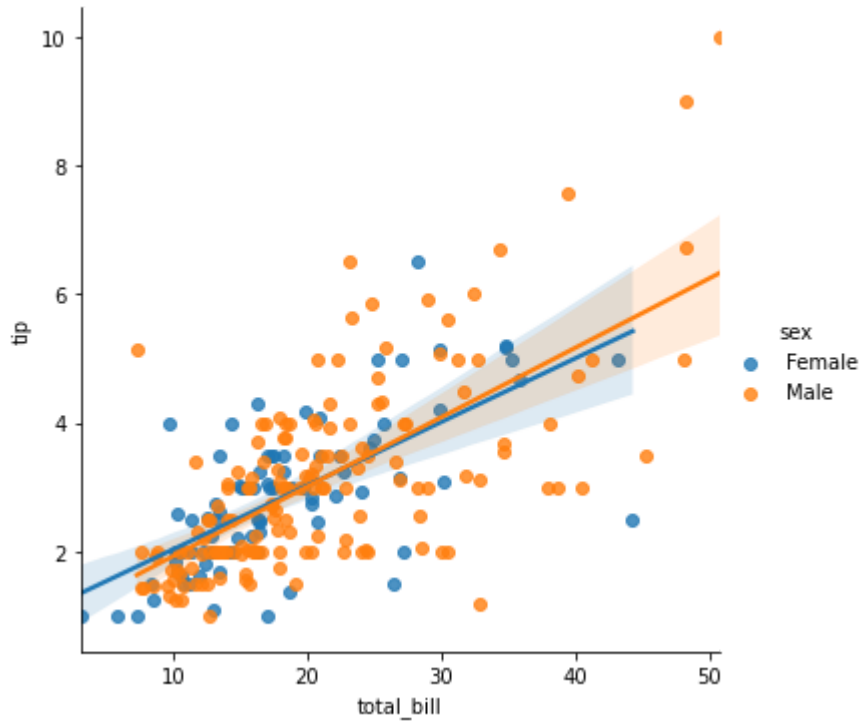
In [65]:

```
ax=plt.subplots()
ax=sns.violinplot(x="time",y="total_bill",data=tips,split=True,hue="sex")
plt.show()
```



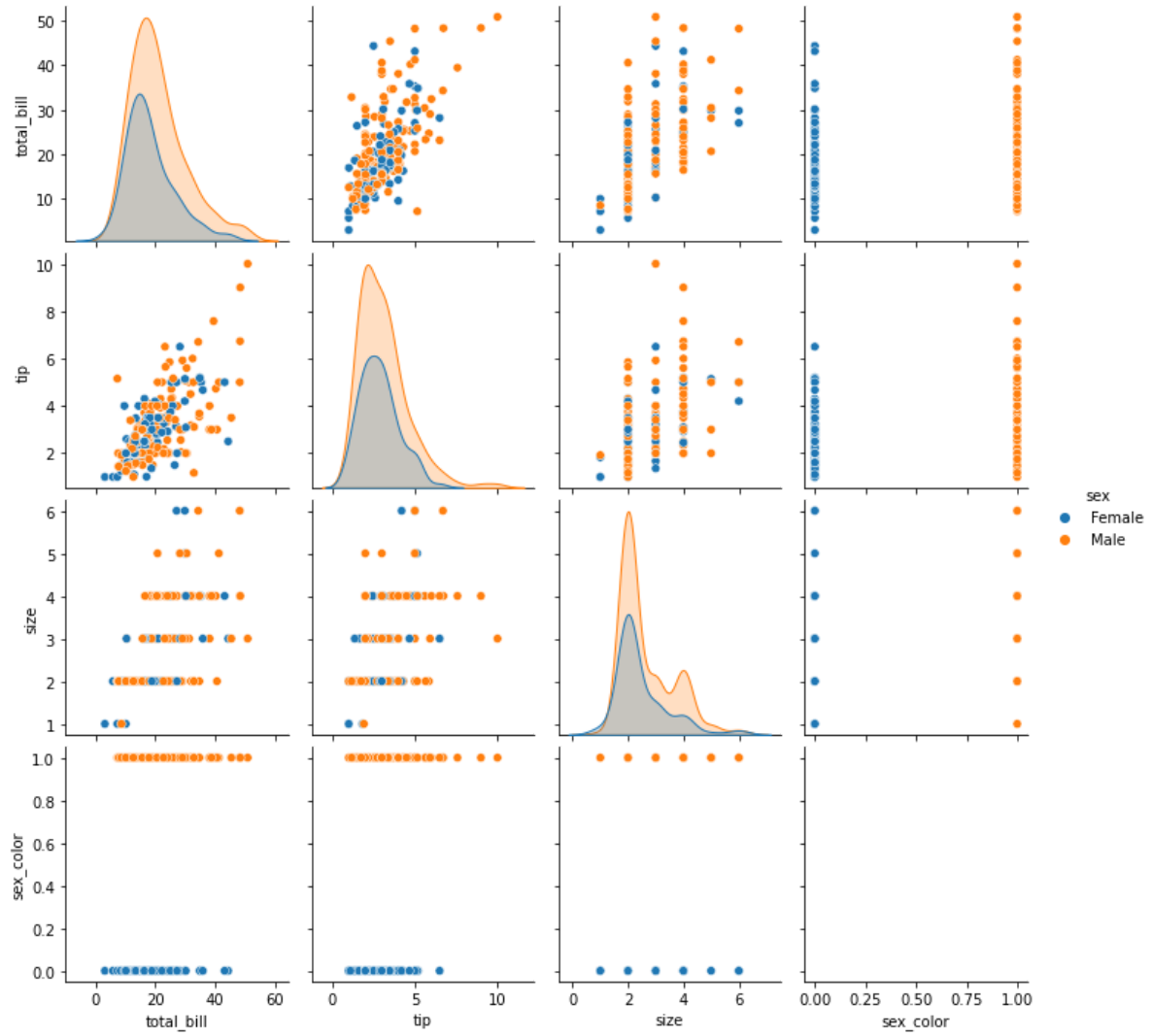
In [66]:

```
sscatter=sns.lmplot(x="total_bill",y="tip",data=tips,hue="sex",fit_reg=True)
plt.show()
```



In [67]:

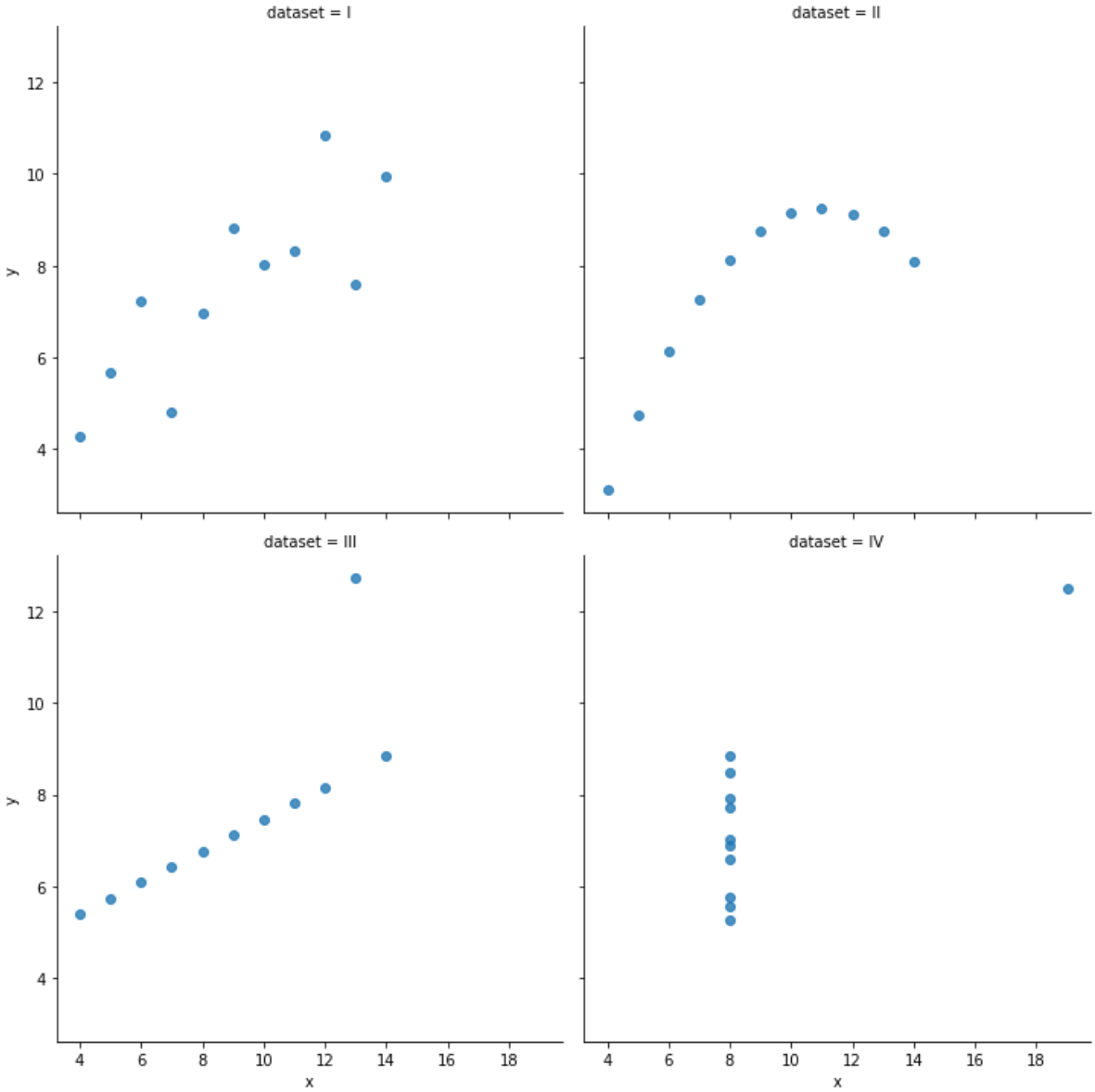
```
fig=sns.pairplot(tips,hue="sex")
plt.show()
```



1.8.13 分面技术

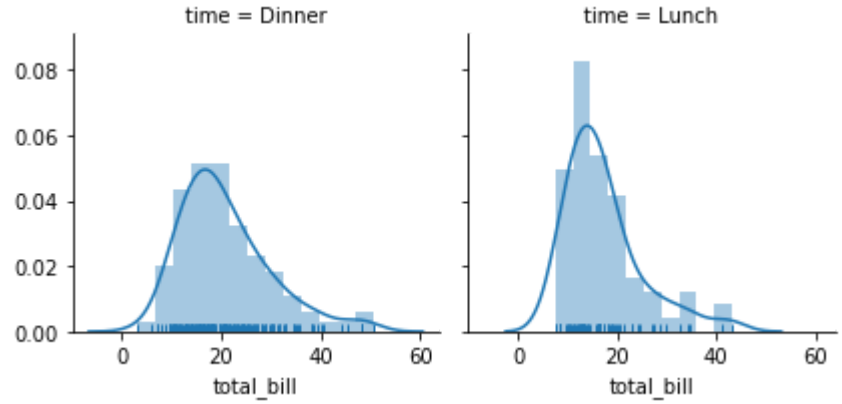
In [68]:

```
# col表示按什么属性分图，col_warp=2意味着每行只有2个图，如果默认的画，4个图在一行。
anscombe_plot=sns.lmplot(x="x",y="y",data=anscombe,fit_reg=False,col="dataset",col_w
```



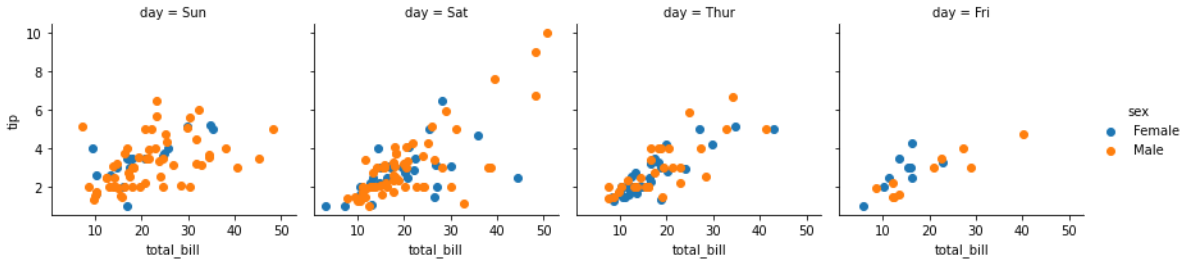
In [69]:

```
# 使用facetGrid进行画图
facet=sns.FacetGrid(tips,col="time")
facet.map(sns.distplot,"total_bill",rug=True)
plt.show()
```



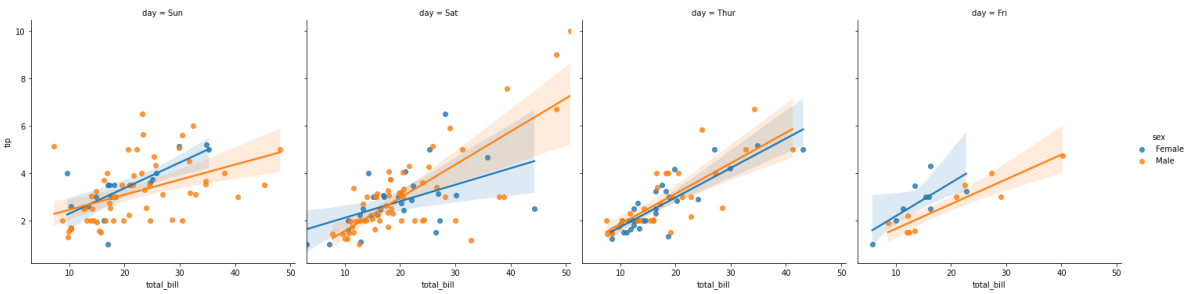
In [70]:

```
# 对于而变量和多变量也可以使用FacetGrid
facet=sns.FacetGrid(tips,col="day",hue="sex")
facet.map(plt.scatter,"total_bill","tip")
facet=facet.add_legend()
plt.show()
```



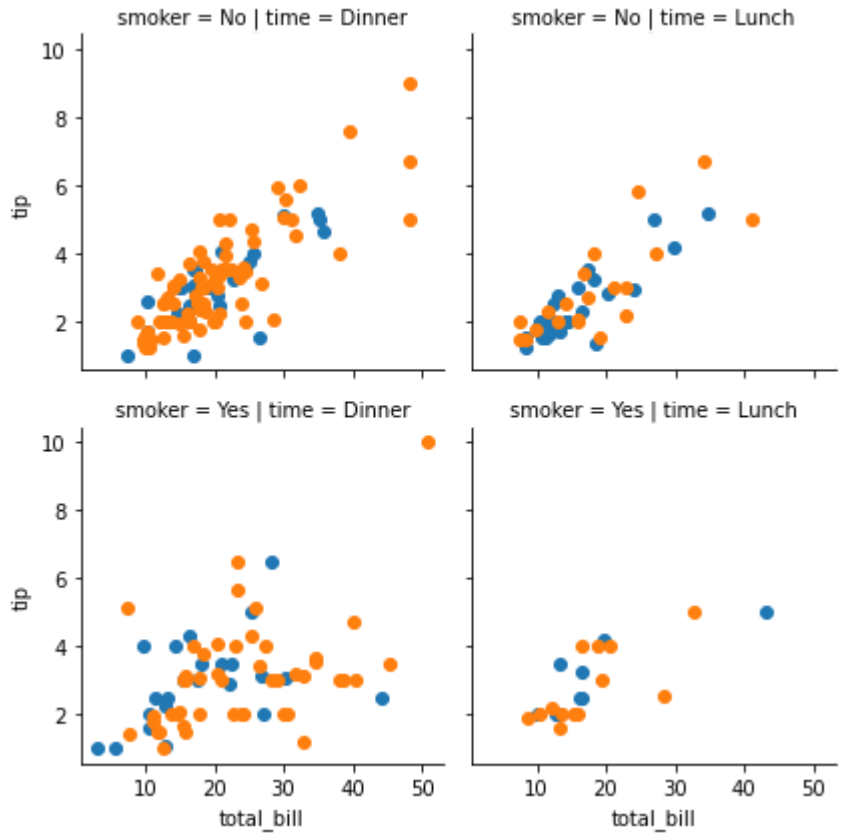
In [71]:

```
# 同样的也可使用seaborn的lmplot
fig=sns.lmplot(x="total_bill",y="tip",hue="sex",data=tips,col="day",fit_reg="True")
```



In [72]:

```
facet=sns.FacetGrid(tips,col="time",row="smoker",hue="sex")
facet.map(plt.scatter,"total_bill","tip")
plt.show()
```



In [73]:

```
facet=sns.factorplot(data=tips,x="day",y="total_bill",hue="sex",row="smoker",col="ti
```