Distribution Plots

localhost:8888/notebooks/Python-Data-Science-and-Machine-Learning-Bootcamp-master/Python-Data-Science-and-Machine-Learning-Bootcamp-master/Python-for-Data-Visualization/Matplotlib/seaborn.ipynb

jupyter seaborn (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help Python [conda env:py35]

In [10]:
```
import seaborn as sns
```

In [11]:
```
%matplotlib inline
```

In [12]:
```
tips = sns.load_dataset('tips')
```

In [13]:
```
tips.head()
```

Out[13]:

	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

localhost:8888/notebooks/Python-Data-Science-and-Machine-Learning-Bootcamp-master/Python-Data-Science-and-Machine-Learning-Bootcamp-master/Python-for-Data-Visualization/Matplotlib/seaborn.ipynb

jupyter seaborn (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help Python [conda env:py35]

In [12]:
```
tips = sns.load_dataset('tips')
```

In [13]:
```
tips.head()
```

Out[13]:

	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

In [ ]:

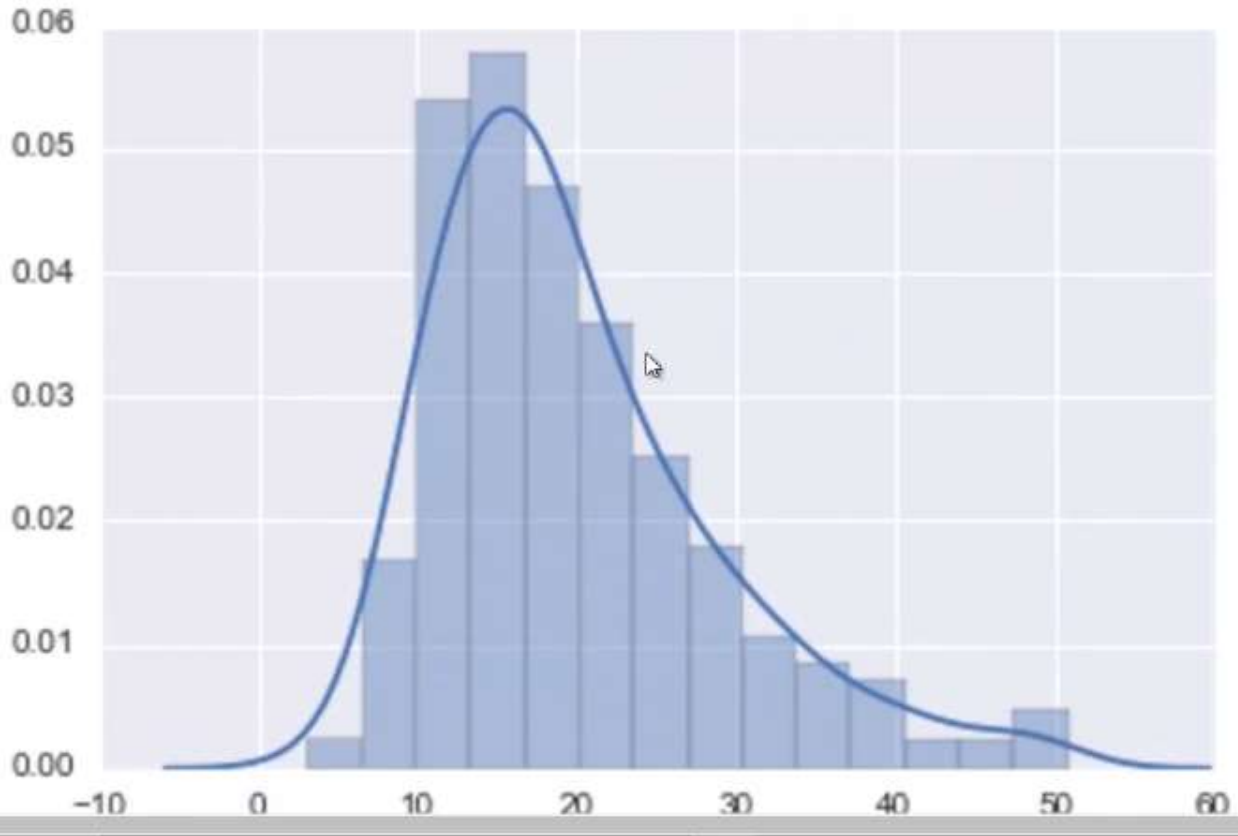
jupyter seaborn (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help

Python [conda env:py35]

In [14]:
```
sns.distplot(tips['total_bill'])
```

Out[14]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x254beb09b00>
```

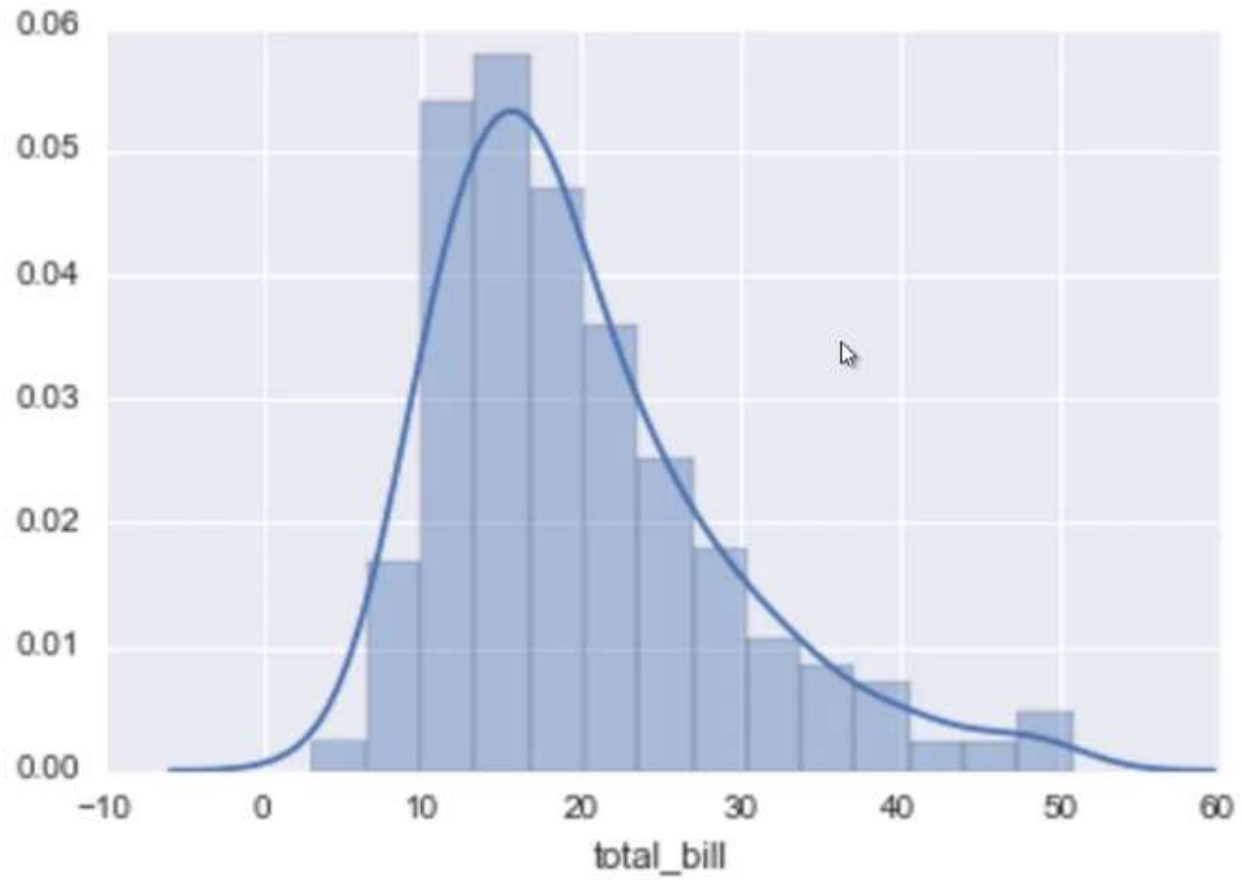


localhost:8888/notebooks/Python-Data-Science-and-Machine-Learning-Bootcamp-master/Python-Data-Science-and-Machine-Learning-Bootcamp-master/Python-for-Data-Visualization/Matplotlib/seaborn.ipynb

jupyter seaborn (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help Python [conda env:py35]

```
Out[14]: <matplotlib.axes._subplots.AxesSubplot at 0x254beb09b00>
```



localhost:8888/notebooks/Python-Data-Science-and-Machine-Learning-Bootcamp-master/Python-Data-Science-and-Machine-Learning-Bootcamp-master/Python-for-Data-Visualization/Matplotlib/seaborn.ipynb

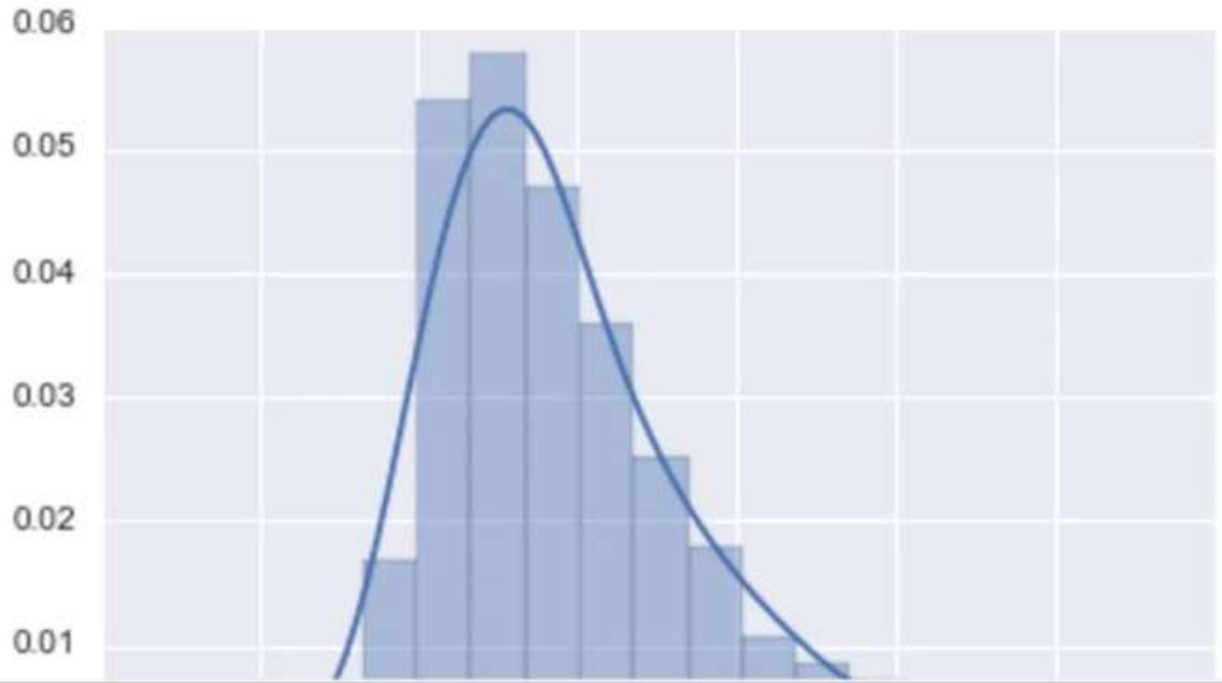
jupyter seaborn (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help Python [conda env:py35]

4	24.59	3.61	Female	No	Sun	Dinner	4

In [14]:
```
sns.distplot(tips['total_bill'],kde=False)
```

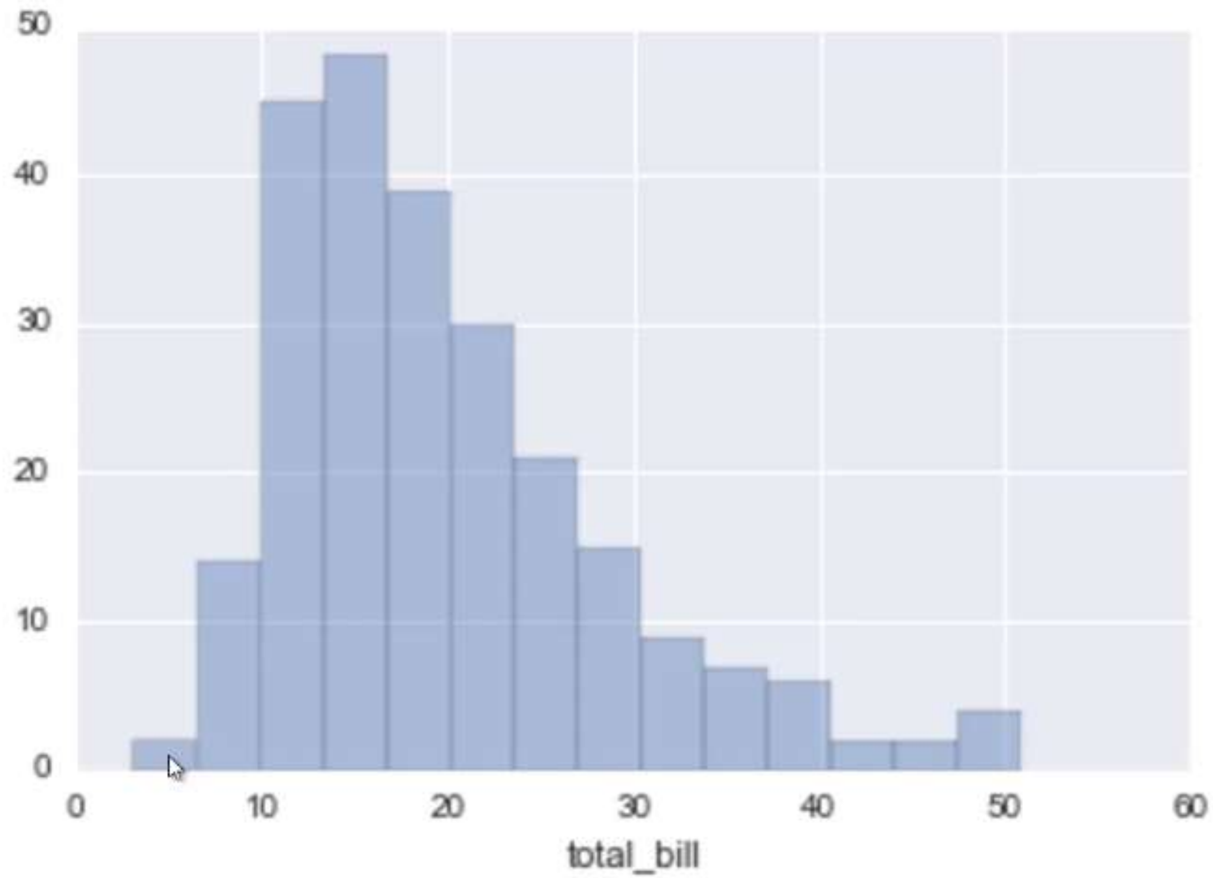
Out[14]: <matplotlib.axes._subplots.AxesSubplot at 0x254beb09b00>



jupyter seaborn (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help Python [conda env:py35]

Out[15]: <matplotlib.axes._subplots.AxesSubplot at 0x254beb3c278>



jupyter seaborn (unsaved changes)

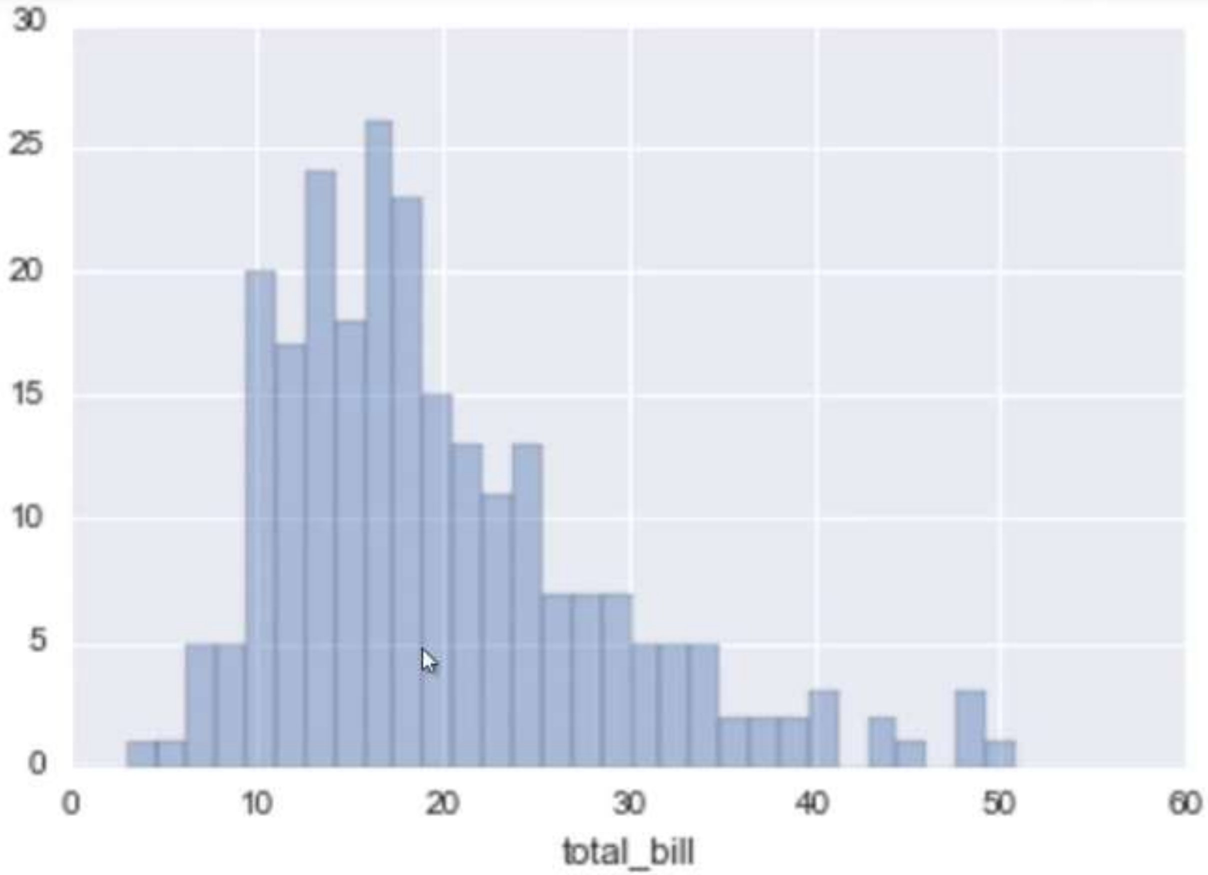
File Edit View Insert Cell Kernel Widgets Help Python [conda env:py35]

In [15]:
```
sns.distplot(tips['total_bill'],kde=False,bins=30)
```

Out[15]: <matplotlib.axes._subplots.AxesSubplot at 0x254beb3c278>

jupyter seaborn (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help | Python [conda env:py35]



In [ ]:

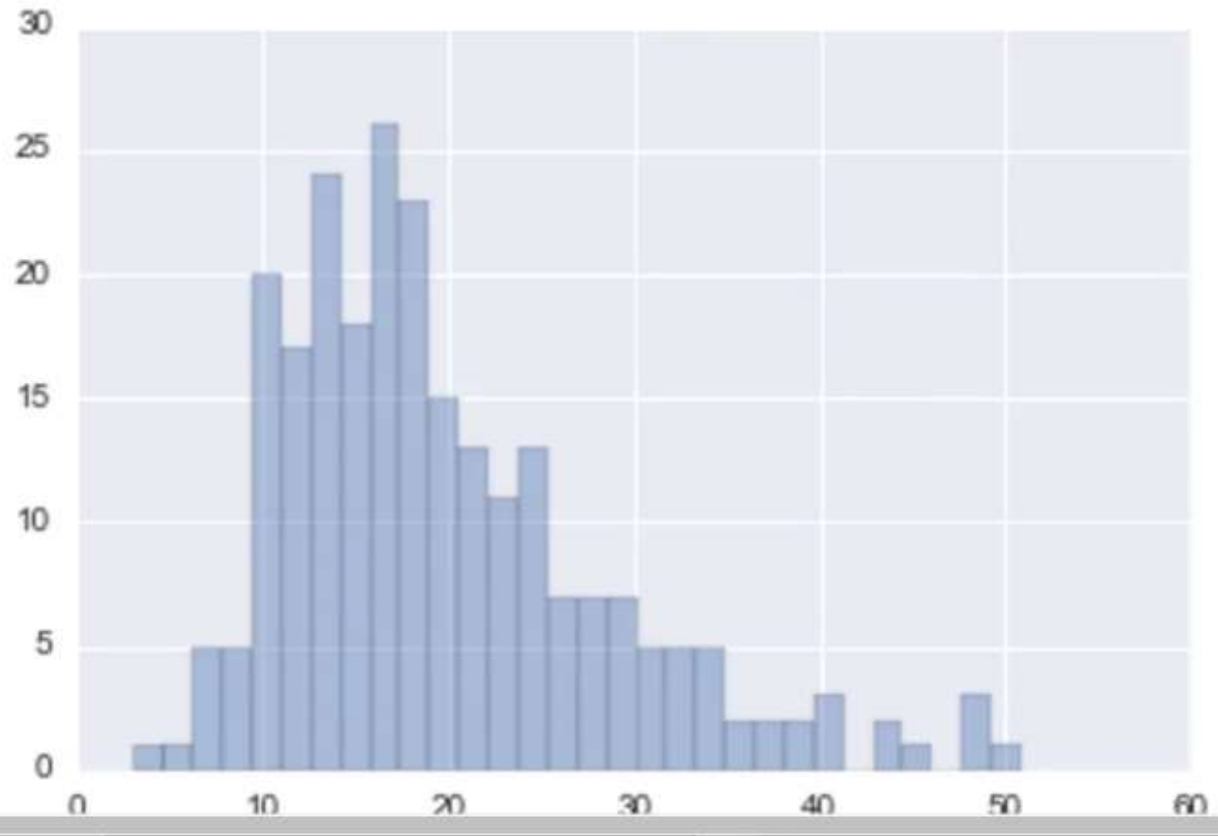
jupyter seaborn (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help

Python [conda env:py35]

In [16]:
```
sns.distplot(tips['total_bill'],kde=False,bins=100)
```

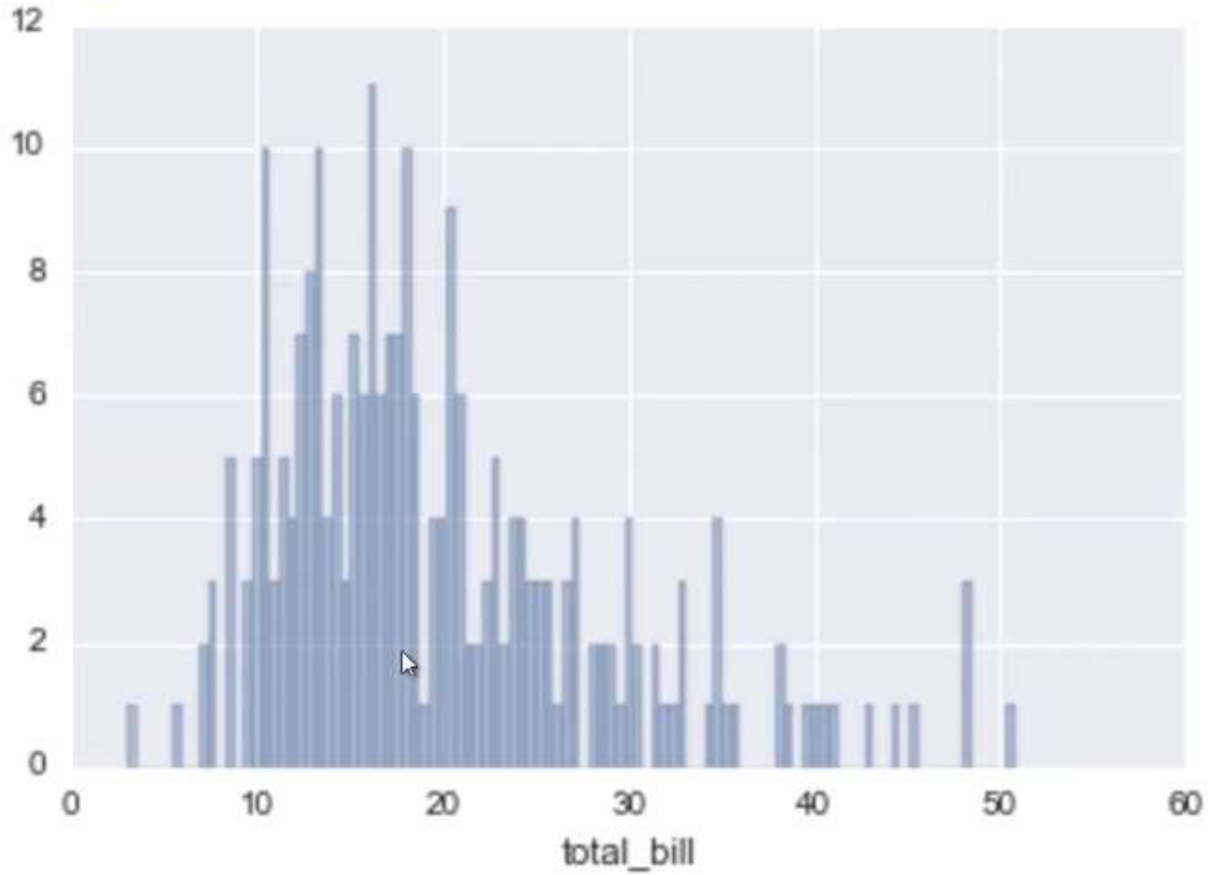
Out[16]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x254bebd97b8>
```



jupyter seaborn (autosaved)

Notebook saved | Python [conda env:py35]

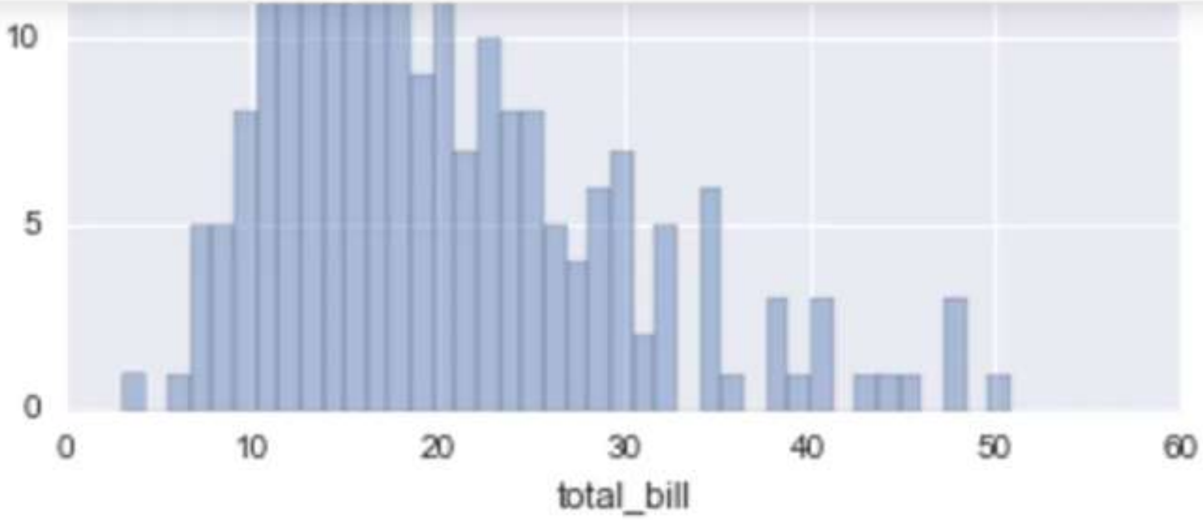
File Edit View Insert Cell Kernel Widgets Help



localhost:8888/notebooks/Python-Data-Science-and-Machine-Learning-Bootcamp-master/Python-Data-Science-and-Machine-Learning-Bootcamp-master/Python-for-Data-Visualization/Matplotlib/seaborn.ipynb

jupyter seaborn (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help Python [conda env:py35]



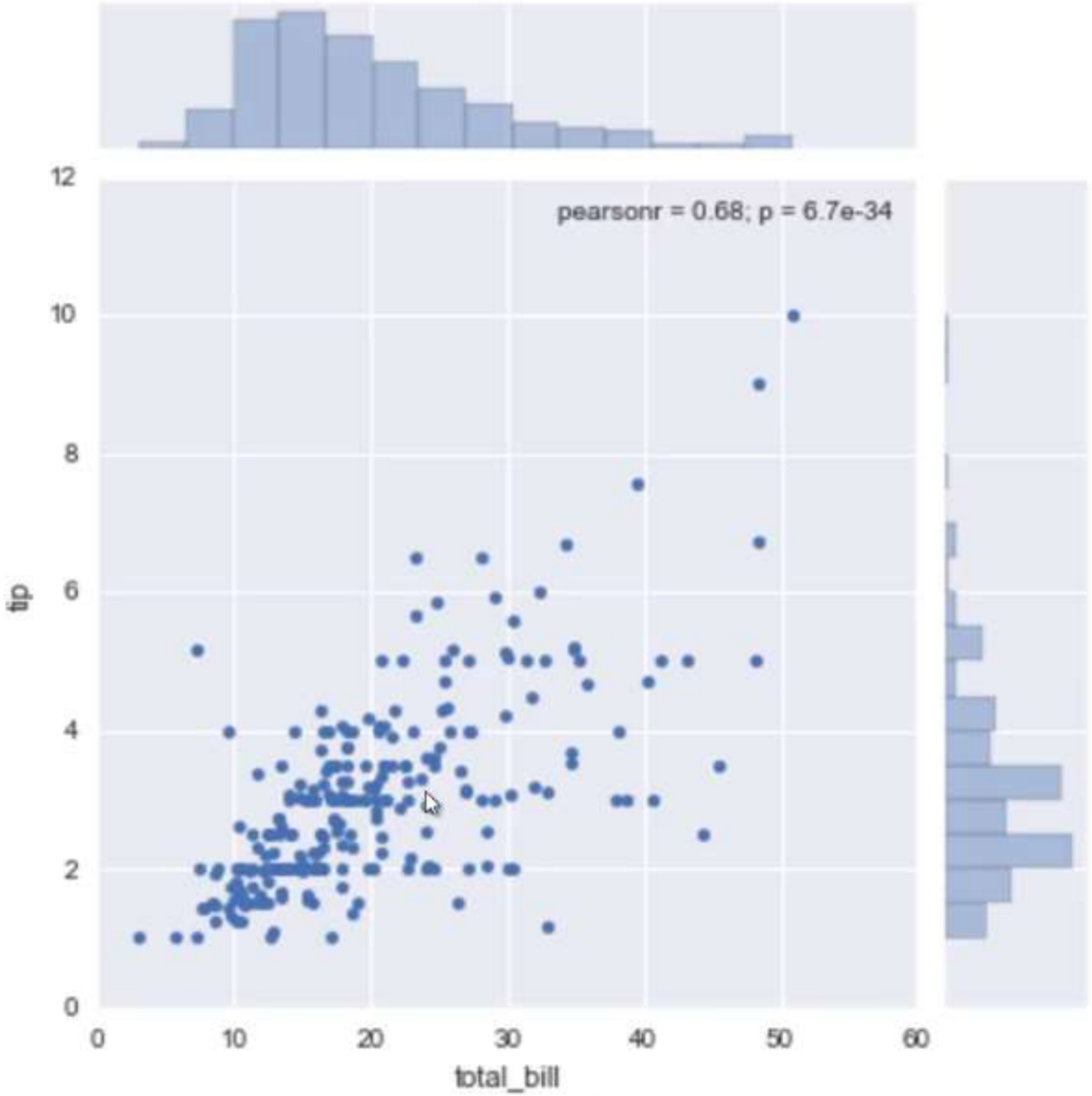

```
In [ ]: sns.jointplot(x='total_bill',y='tip',data=tips)
```

To do Bi-variate data, it is useful.

localhost:8888/notebooks/Python-Data-Science-and-Machine-Learning-Bootcamp-master/Python-Data-Science-and-Machine-Learning-Bootcamp-master/Python-for-Data-Visualization/Matplotlib/seaborn.ipynb

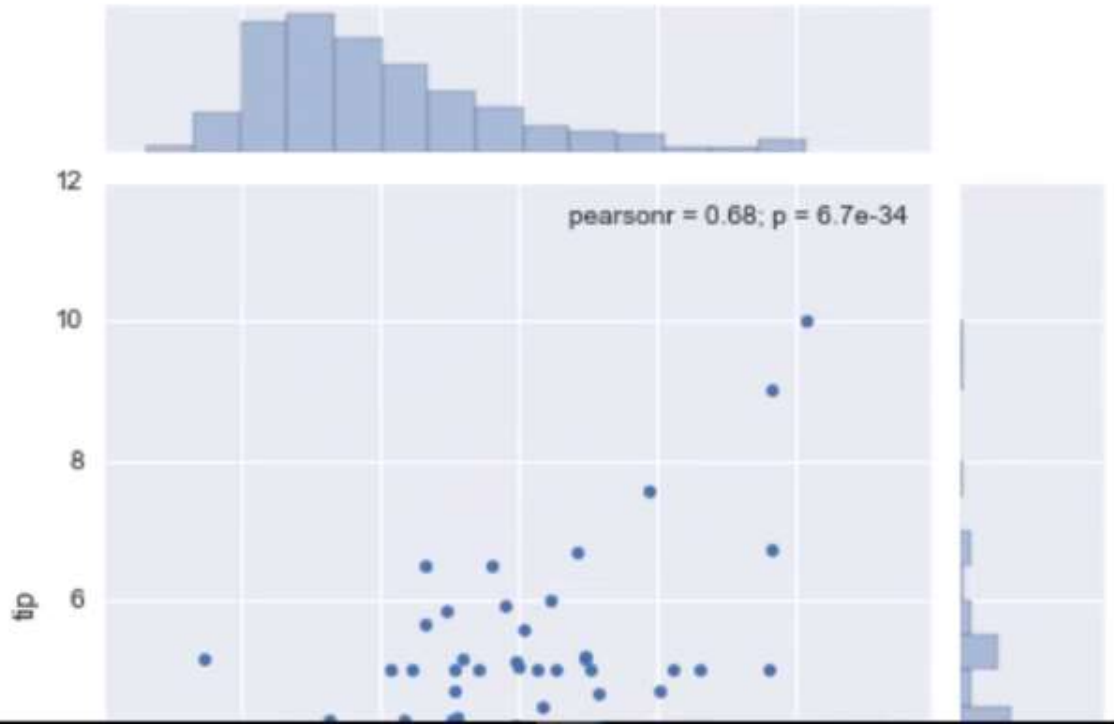
jupyter seaborn (autosaved)

File Edit View Insert Cell Kernel Widgets Help Python [conda env:py35]



In [19]:
```
sns.jointplot(x='total_bill',y='tip',data=tips,kind='hex')
```

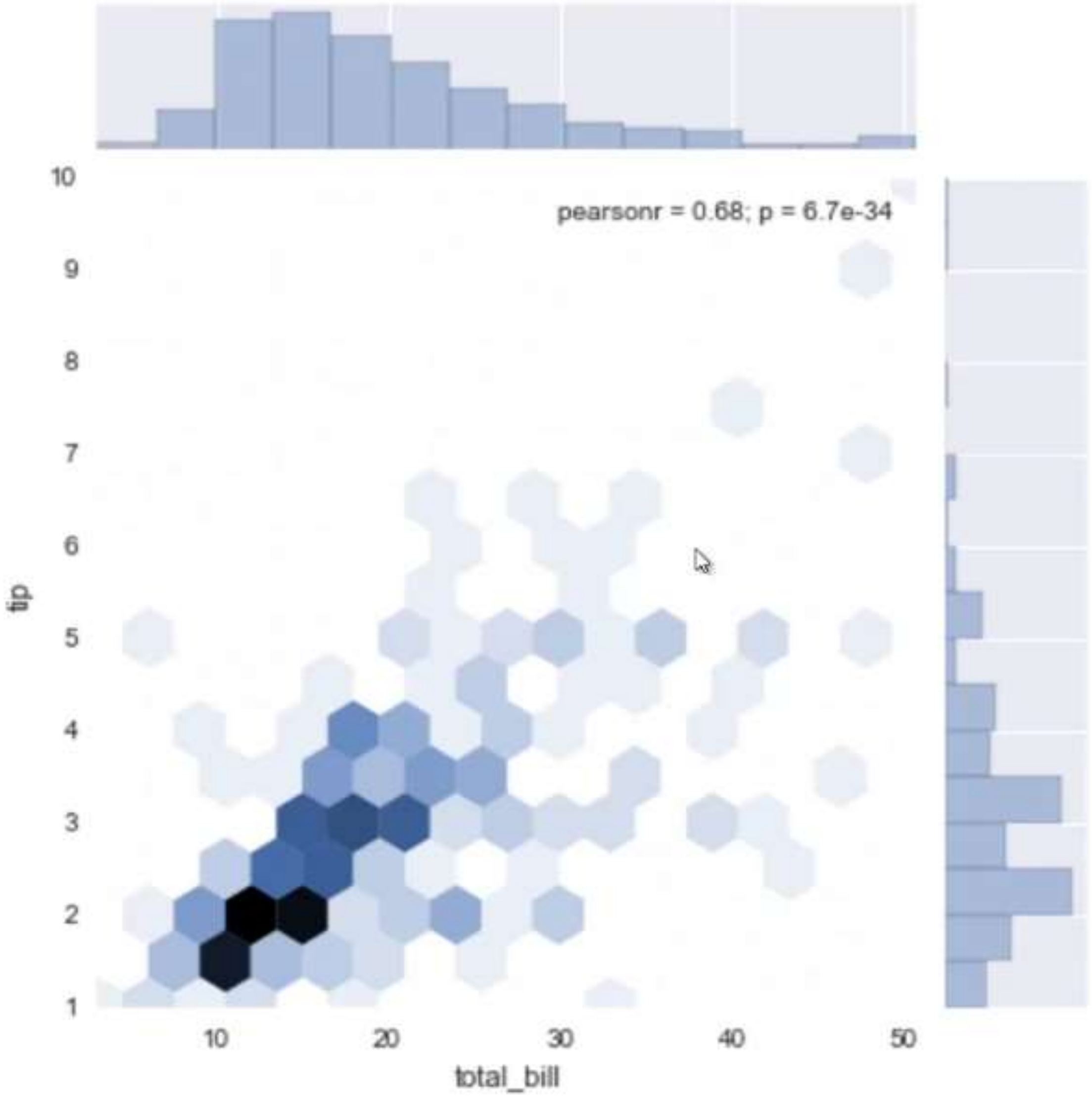
Out[19]: <seaborn.axisgrid.JointGrid at 0x254befa8a90>



localhost:8888/notebooks/Python-Data-Science-and-Machine-Learning-Bootcamp-master/Python-Data-Science-and-Machine-Learning-Bootcamp-master/Python-for-Data-Visualization/Matplotlib/seaborn.ipynb

jupyter seaborn (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help | Python [conda env:py35]



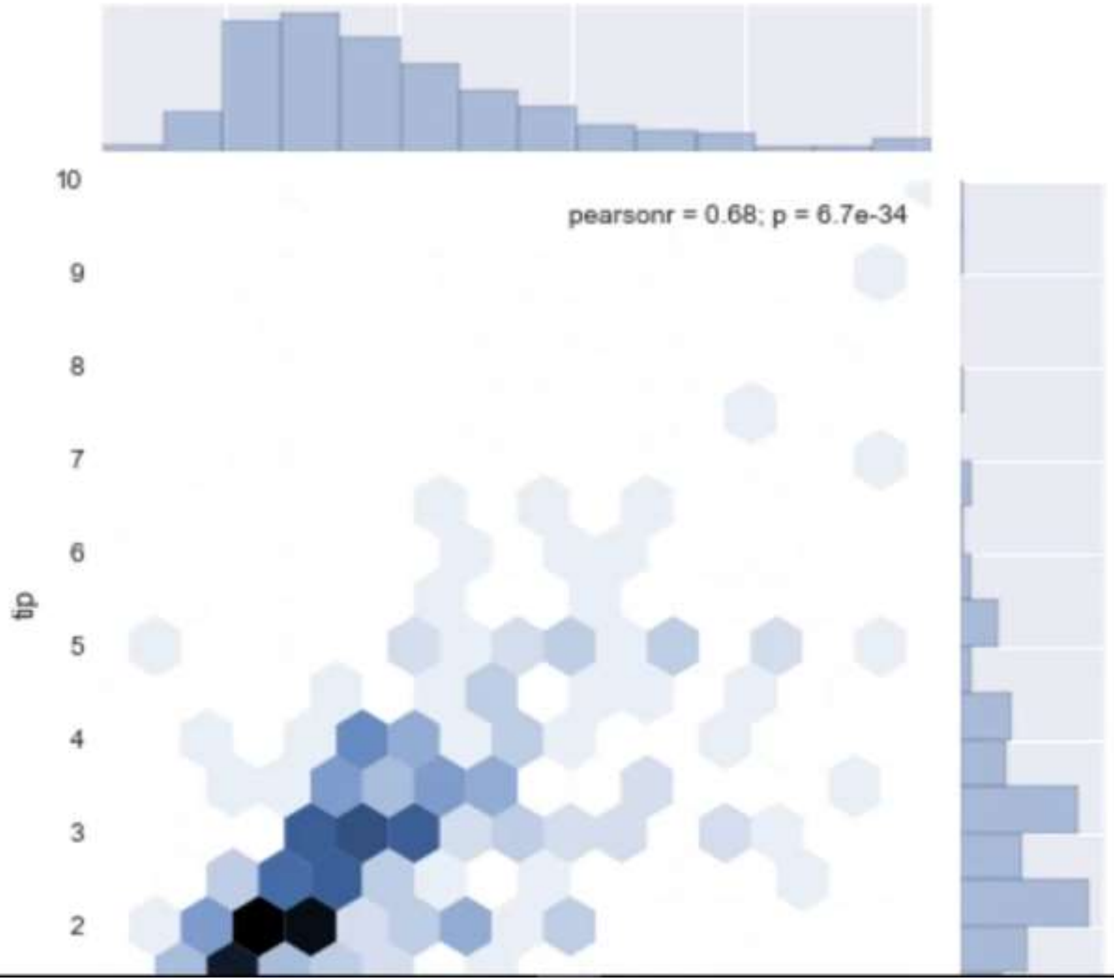
localhost:8888/notebooks/Python-Data-Science-and-Machine-Learning-Bootcamp-master/Python-Data-Science-and-Machine-Learning-Bootcamp-master/Python-for-Data-Visualization/Matplotlib/seaborn.ipynb

jupyter seaborn (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help Python [conda env:py35]

In [20]:
```
sns.jointplot(x='total_bill',y='tip',data=tips,kind='reg')
```

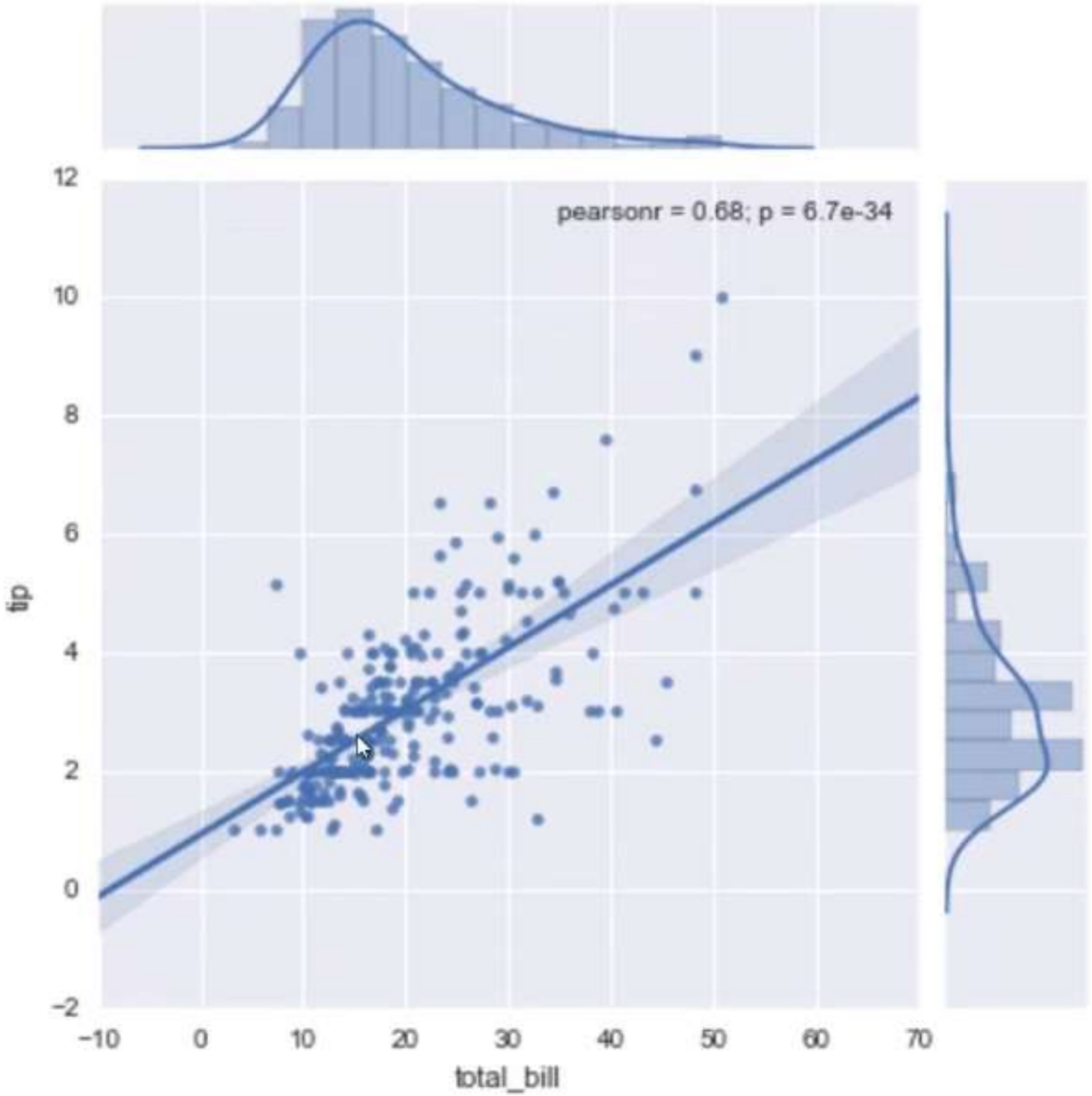
Out[20]: <seaborn.axisgrid.JointGrid at 0x254bf0d6780>



localhost:8888/notebooks/Python-Data-Science-and-Machine-Learning-Bootcamp-master/Python-Data-Science-and-Machine-Learning-Bootcamp-master/Python-for-Data-Visualization/Matplotlib/seaborn.ipynb

jupyter seaborn (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help | Python [conda env:py35]



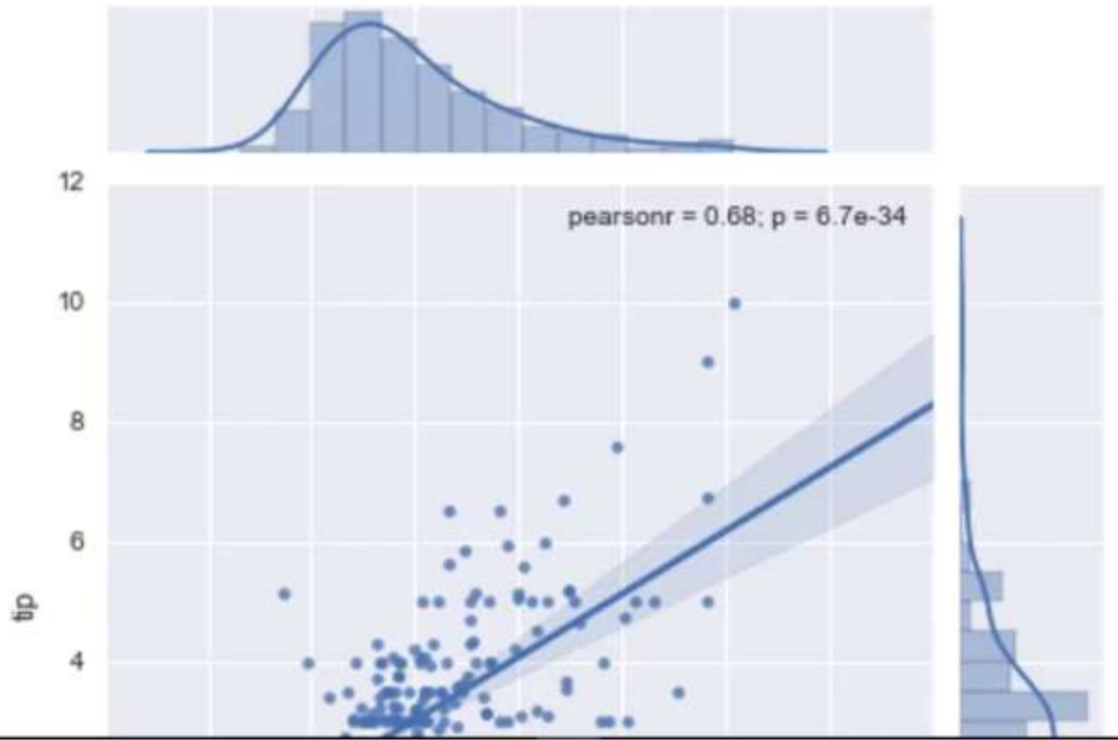
jupyter seaborn (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help

Python [conda env:py35]

In [21]:
```
sns.jointplot(x='total_bill',y='tip',data=tips,kind='kde')
```

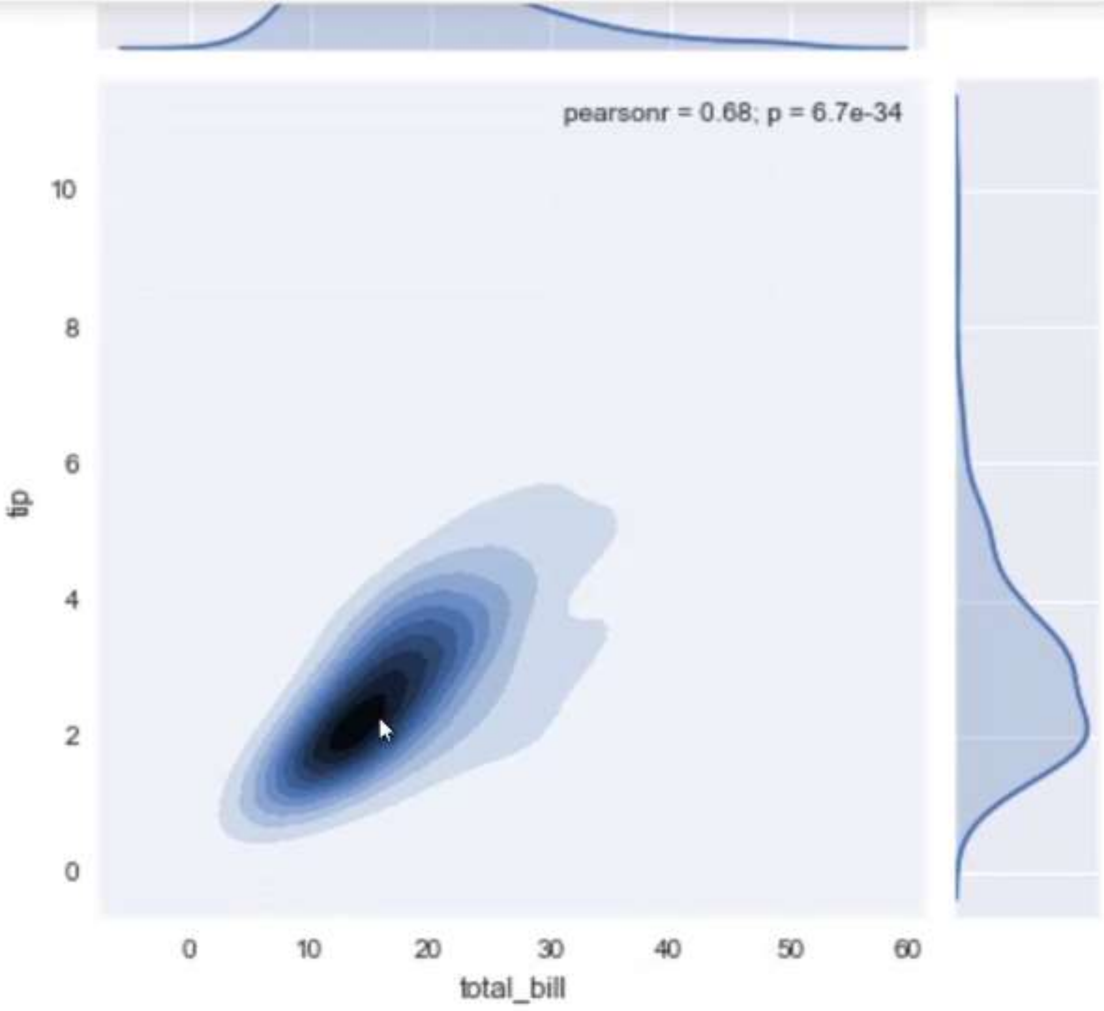
Out[21]: <seaborn.axisgrid.JointGrid at 0x254bf2c5898>



localhost:8888/notebooks/Python-Data-Science-and-Machine-Learning-Bootcamp-master/Python-Data-Science-and-Machine-Learning-Bootcamp-master/Python-for-Data-Visualization/Matplotlib/seaborn.ipynb

jupyter seaborn (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help Python [conda env:py35]

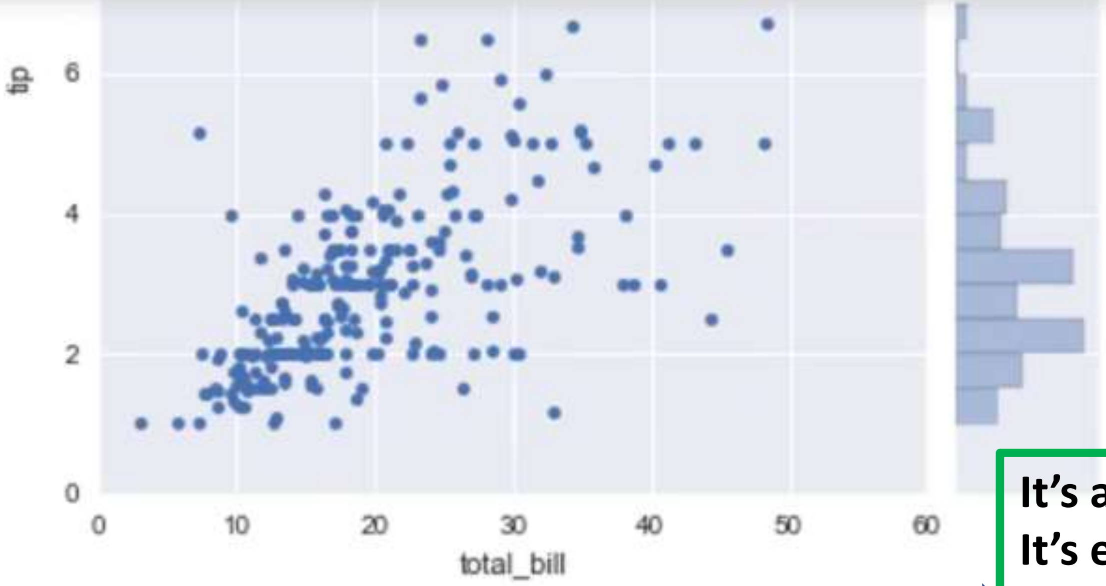


In [ ]:

jupyter seaborn (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help

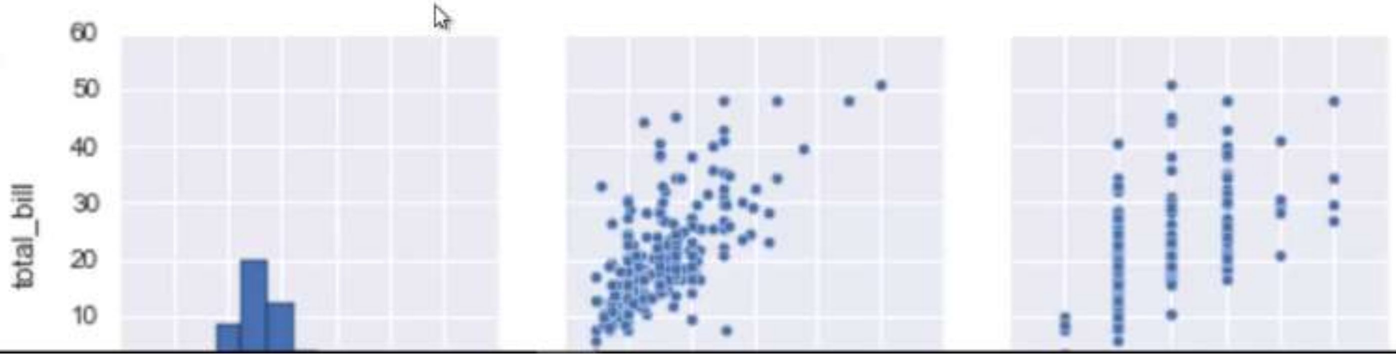
Python [conda env:py35]



It’s a building up on the joint plot. It’s essentially a joint plot for every single column or numerical column for the data set.

In [24]:
```
sns.pairplot(tips)
```

Out[24]:
```
<seaborn.axisgrid.PairGrid at 0x254bebdde10>
```



Python-for-Data-Visualiz × seaborn ×

localhost:8888/notebooks/Python-Data-Science-and-Machine-Learning-Bootcamp-master/Python-Data-Science-and-Machine-Learning-Bootcamp-master/Python-for-Data-Visualization/Matplotlib/seaborn.ipynb

jupyter seaborn (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help

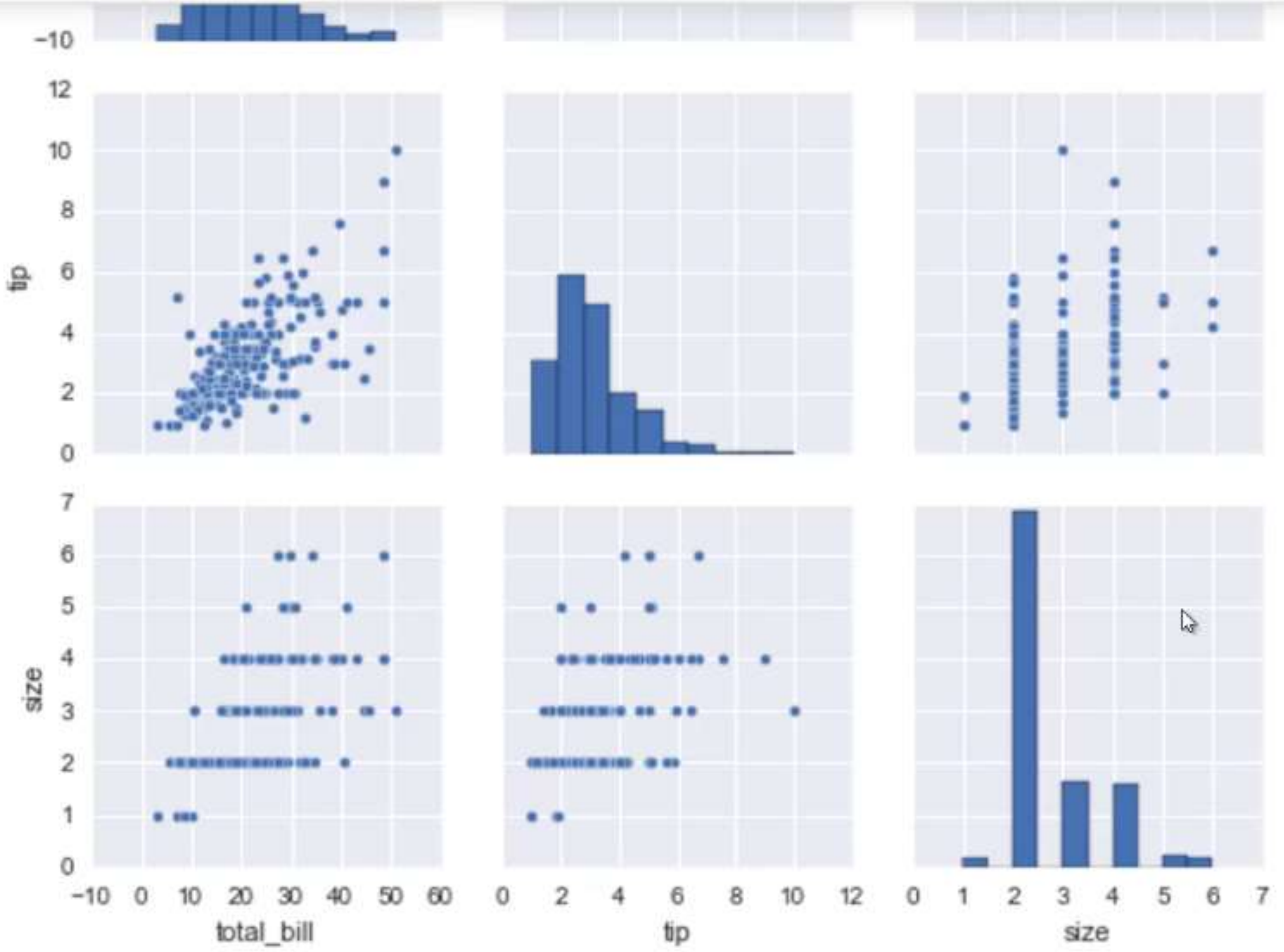
Python [conda env:py35]

Python-for-Data-Visuali: × seaborn ×

localhost:8888/notebooks/Python-Data-Science-and-Machine-Learning-Bootcamp-master/Python-Data-Science-and-Machine-Learning-Bootcamp-master/Python-for-Data-Visualization/Matplotlib/seaborn.ipynb

jupyter seaborn (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help Python [conda env:py35]



In [ ]:

jupyter seaborn (unsaved changes)

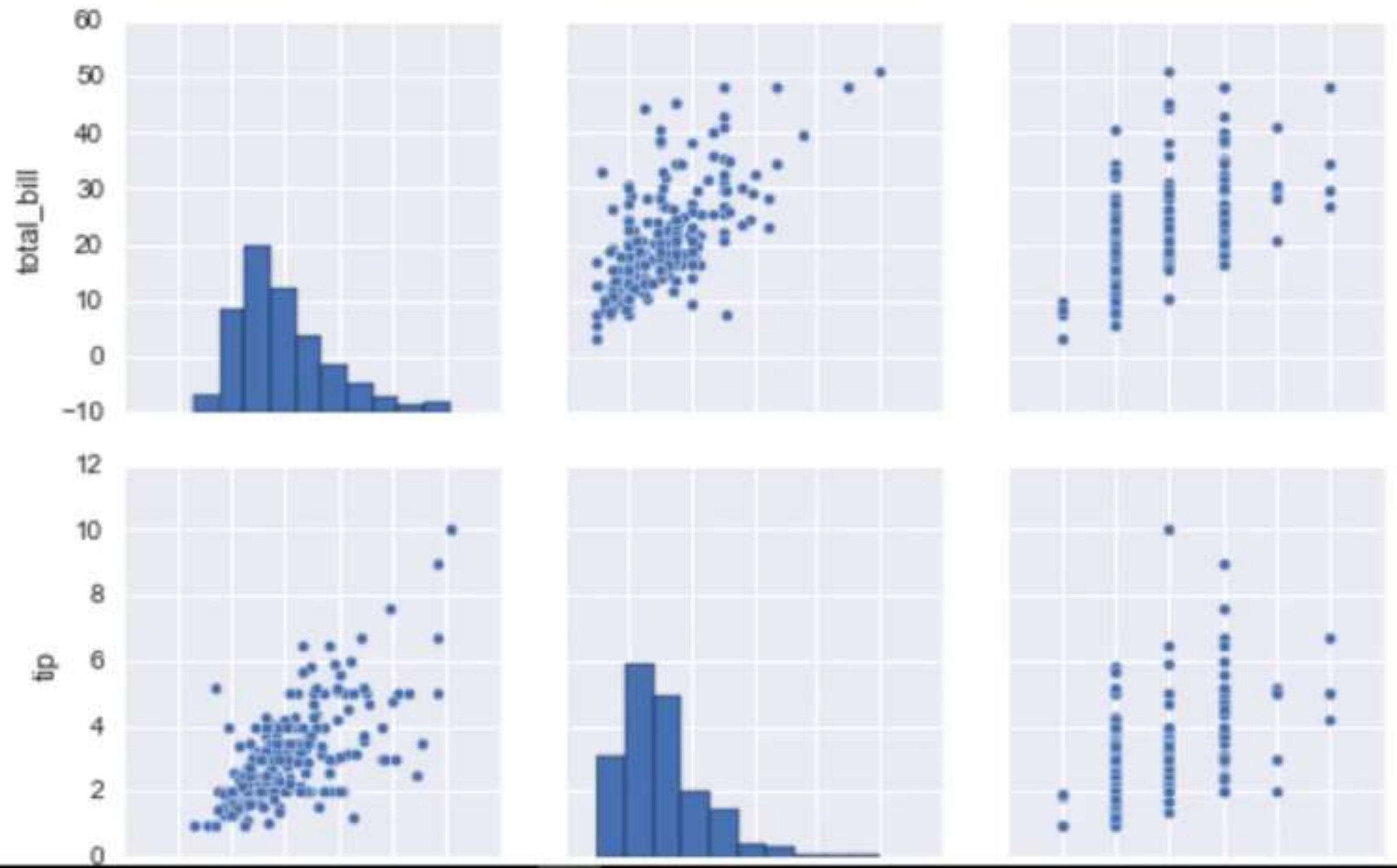
File Edit View Insert Cell Kernel Widgets Help Python [conda env:py35]

total_bill

In [24]:
```
sns.pairplot(tips,hue='sex')
```

Only categorical; no numerical is allowed

Out[24]: <seaborn.axisgrid.PairGrid at 0x254bebdde10>



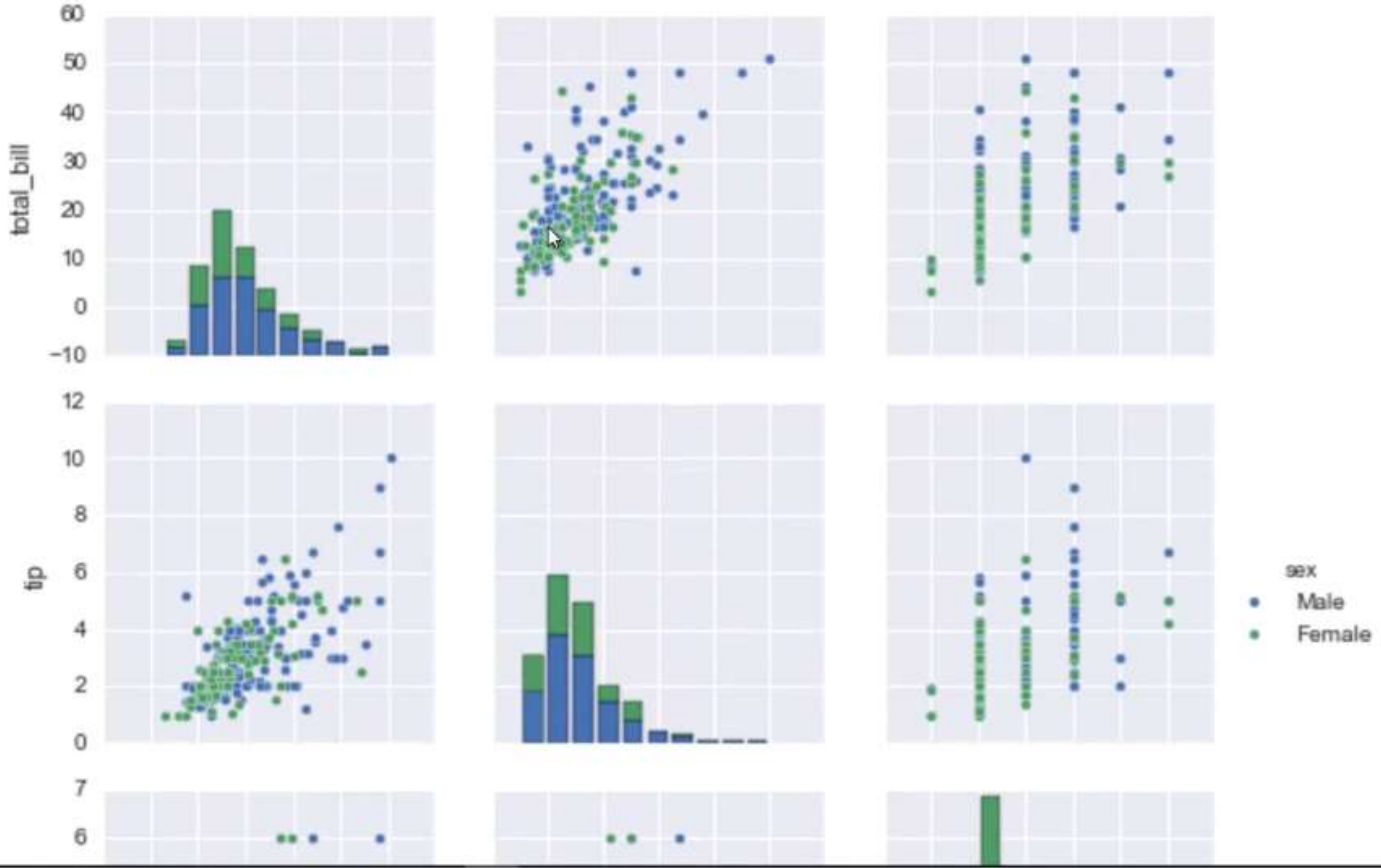
localhost:8888/notebooks/Python-Data-Science-and-Machine-Learning-Bootcamp-master/Python-Data-Science-and-Machine-Learning-Bootcamp-master/Python-for-Data-Visualization/Matplotlib/seaborn.ipynb

jupyter seaborn (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help Python [conda env:py35]

In [25]:
```
sns.pairplot(tips,hue='sex')
```

Out[25]: <seaborn.axisgrid.PairGrid at 0x254be7a5f60>



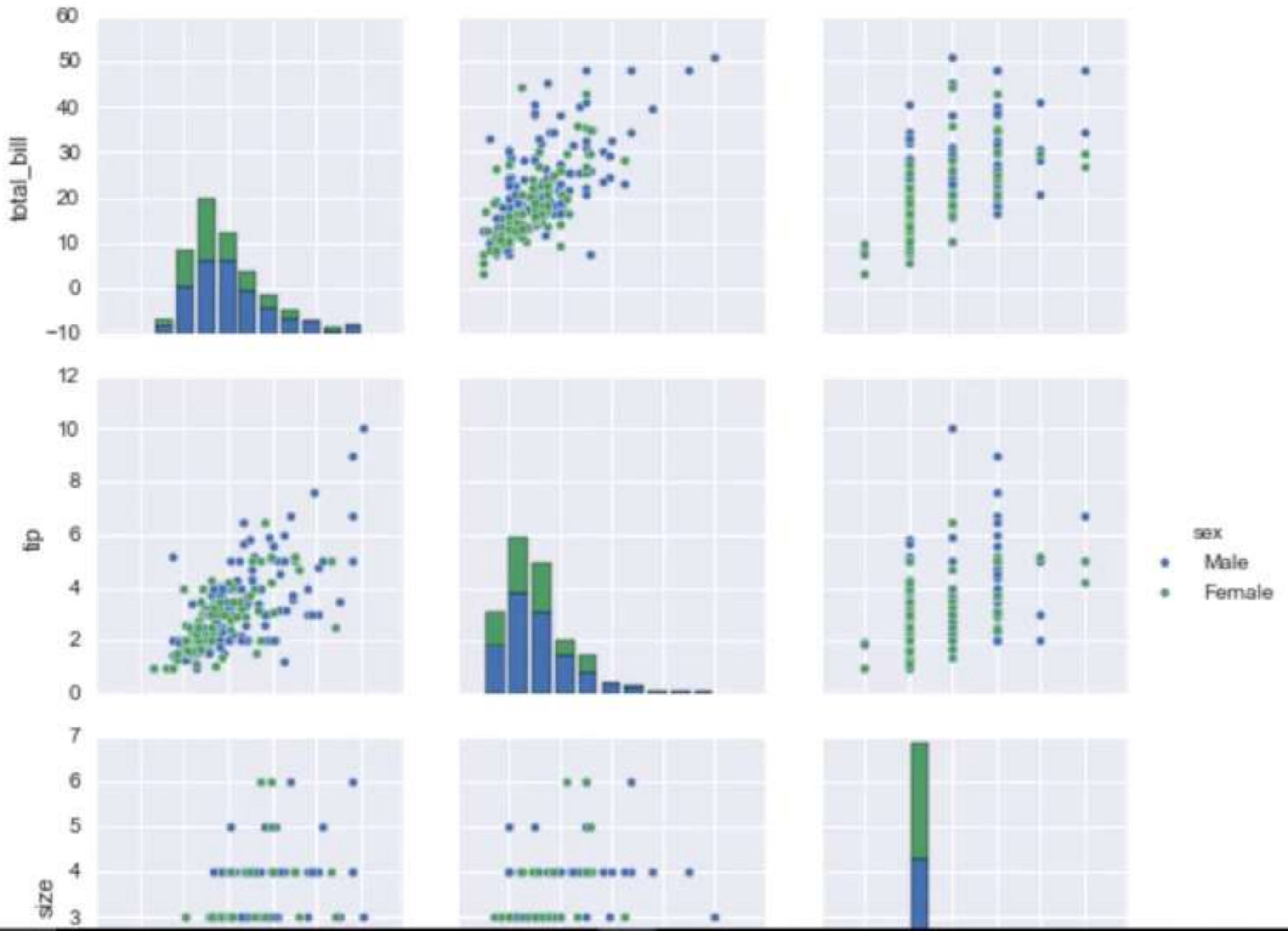
jupyter seaborn Last Checkpoint: 17 minutes ago (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help

Python [conda env:py35]

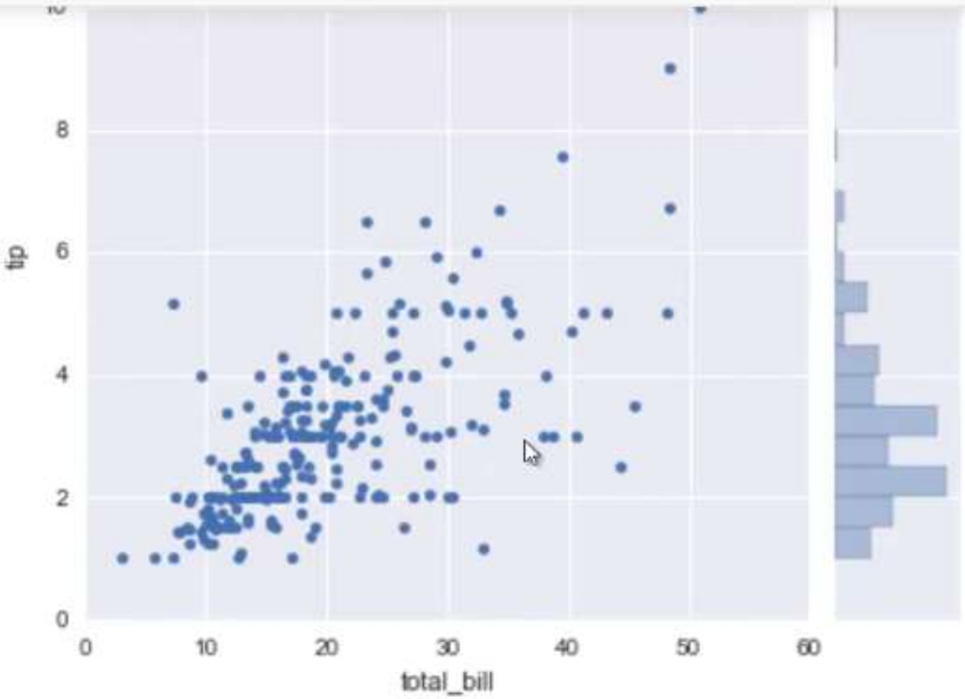
In [25]:
```
sns.pairplot(tips,hue='sex')
```

Out[25]: <seaborn.axisgrid.PairGrid at 0x254be7a5f60>



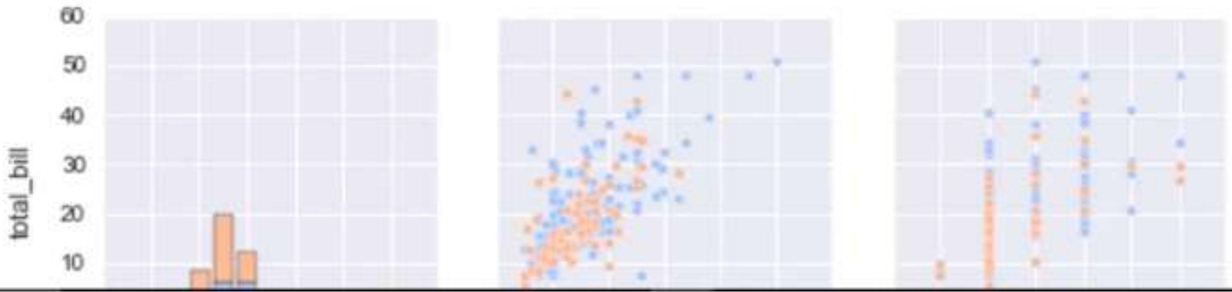
jupyter seaborn Last Checkpoint: 17 minutes ago (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help Python [conda env:py35]



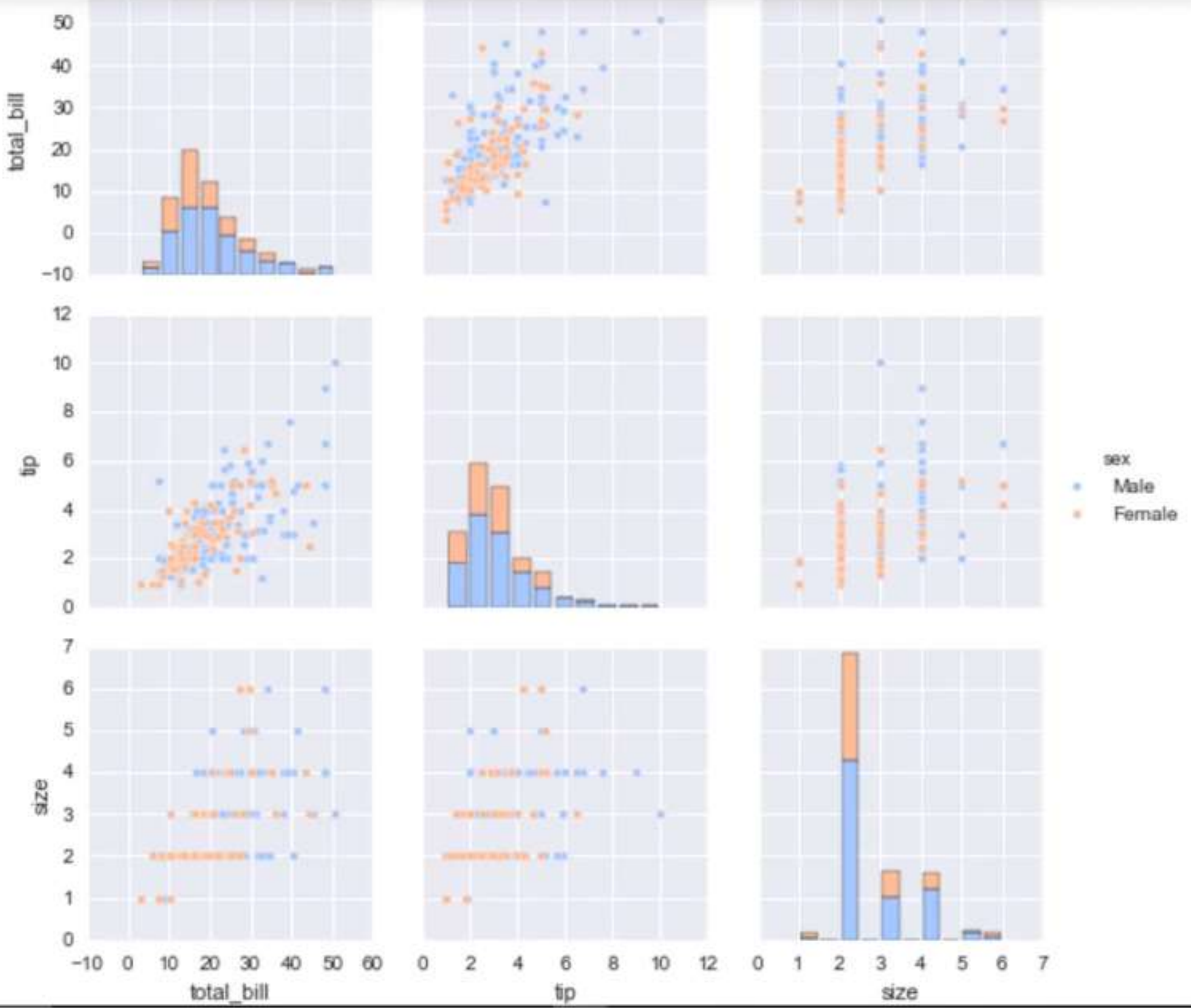
In [26]:
```
sns.pairplot(tips,hue='sex',palette='coolwarm')
```

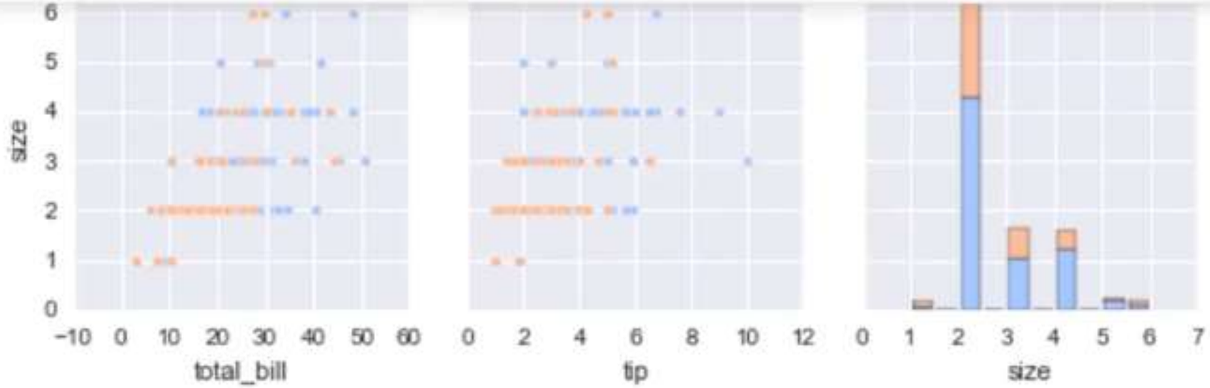
Out[26]: <seaborn.axisgrid.PairGrid at 0x254c24b4940>



jupyter seaborn Last Checkpoint: 17 minutes ago (unsaved changes)

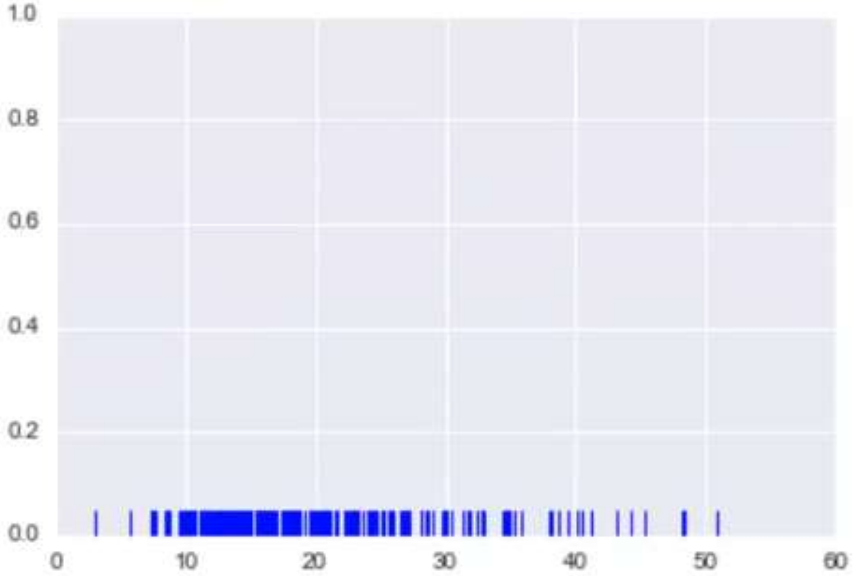
File Edit View Insert Cell Kernel Widgets Help | Python [conda env:py35]





In [27]:
```python
sns.rugplot(tips['total_bill'])
```

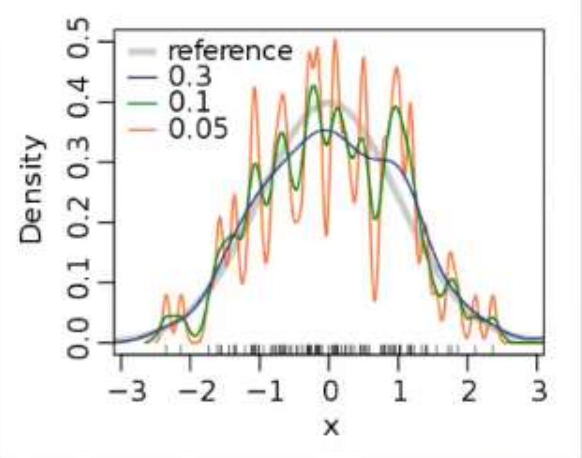
Out[27]: <matplotlib.axes._subplots.AxesSubplot at 0x254c2c8b550>



Kernel density estimation

From Wikipedia, the free encyclopedia

In statistics, **kernel density estimation (KDE)** is a non-parametric way to estimate the probability density function of a random variable. Kernel density estimation is a fundamental data smoothing problem where inferences about the population are made, based on a finite data sample. In some fields such as signal processing and econometrics it is also termed the *Parzen–Rosenblatt window* method, after Emanuel Parzen and Murray Rosenblatt, who are usually credited with independently creating it in its current form.[1][2]

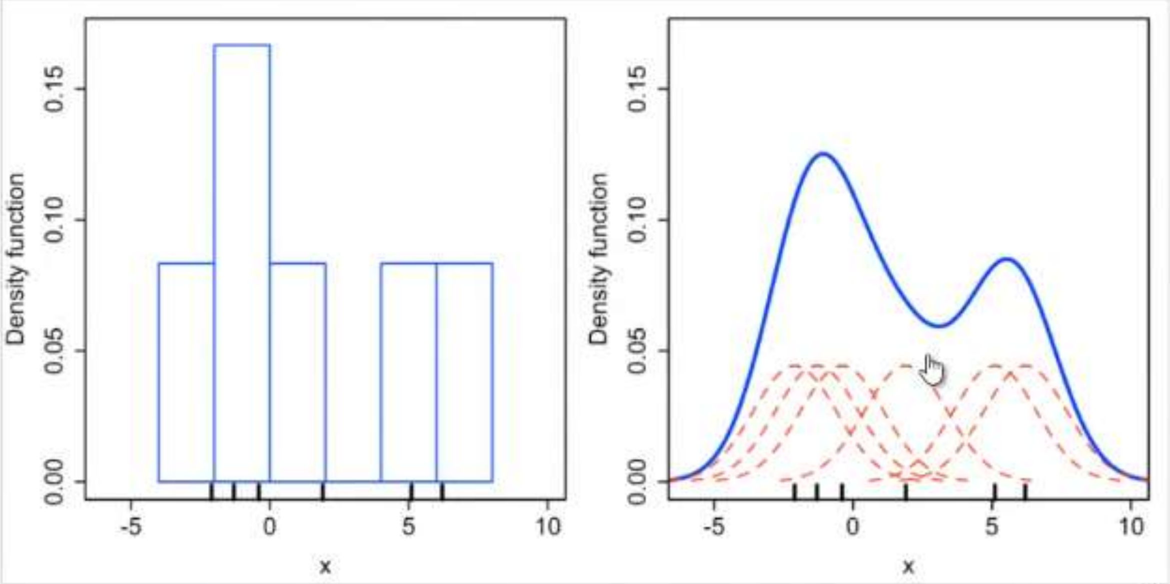


Kernel density estimation of 100 normally distributed random numbers using different smoothing bandwidths.

Contents [hide]

the range of the data. In this case, we have 6 bins each of width 2. Whenever a data point falls inside this interval, we place a box of height 1/12. If more than one data point falls inside the same bin, we stack the boxes on top of each other.

For the kernel density estimate, we place a normal kernel with variance 2.25 (indicated by the red dashed lines) on each of the data points x_i. The kernels are summed to make the kernel density estimate (solid blue curve). The smoothness of the kernel density estimate is evident compared to the discreteness of the histogram, as kernel density estimates converge faster to the true underlying density for continuous random variables.[6]



Comparison of the histogram (left) and kernel density estimate (right) constructed using the same data. The 6 individual kernels are the red dashed curves, the kernel density estimate the blue curves. The data points are the rug plot on the horizontal axis.

Bandwidth selection [edit]

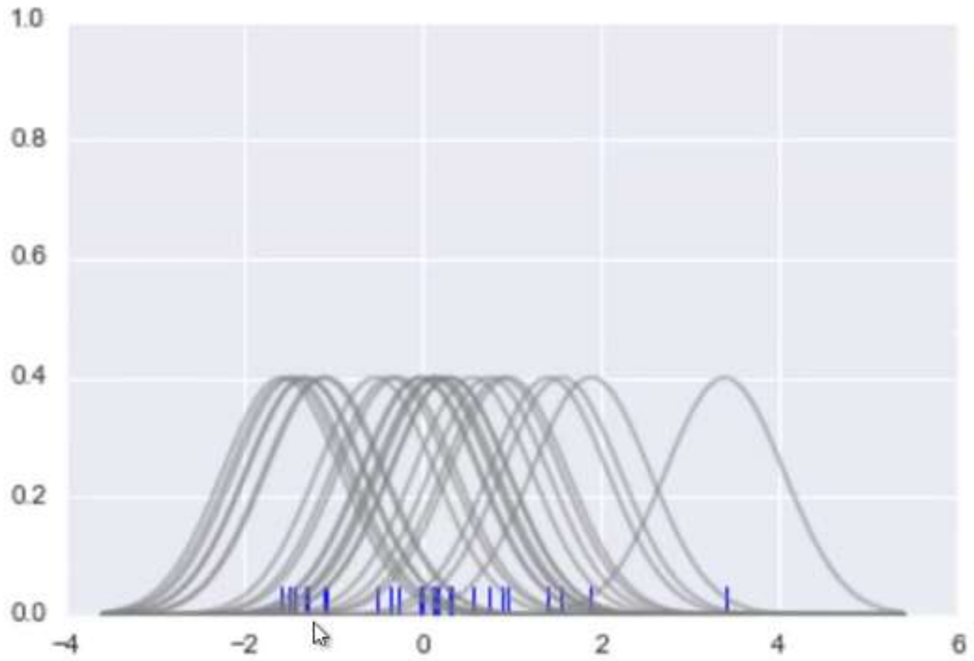
The bandwidth of the kernel is a free parameter which exhibits a strong influence on the resulting estimate. To illustrate its effect, we take a simulated random sample from the standard normal distribution (plotted at the blue spikes in the rug plot on the horizontal axis). The grey curve is the

localhost:8888/notebooks/Python-Data-Science-and-Machine-Learning-Bootcamp-master/Python-Data-Science-and-Machine-Learning-Bootcamp-master/Python-for-Data-Visualization/Matplotlib/seaborn.ipynb

jupyter seaborn (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help

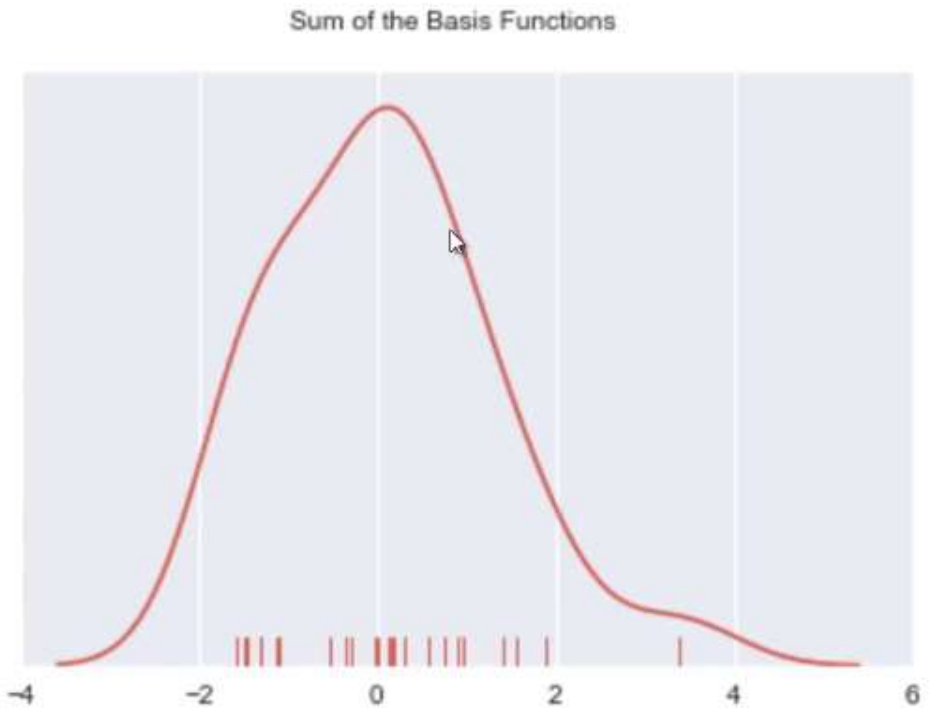
Python [conda env:py35]



In [ ]:

```
plt.suptitle("Sum of the Basis Functions")
```

Out[33]: <matplotlib.text.Text at 0x254c32d7f28>



In [ ]:

jupyter seaborn (unsaved changes)

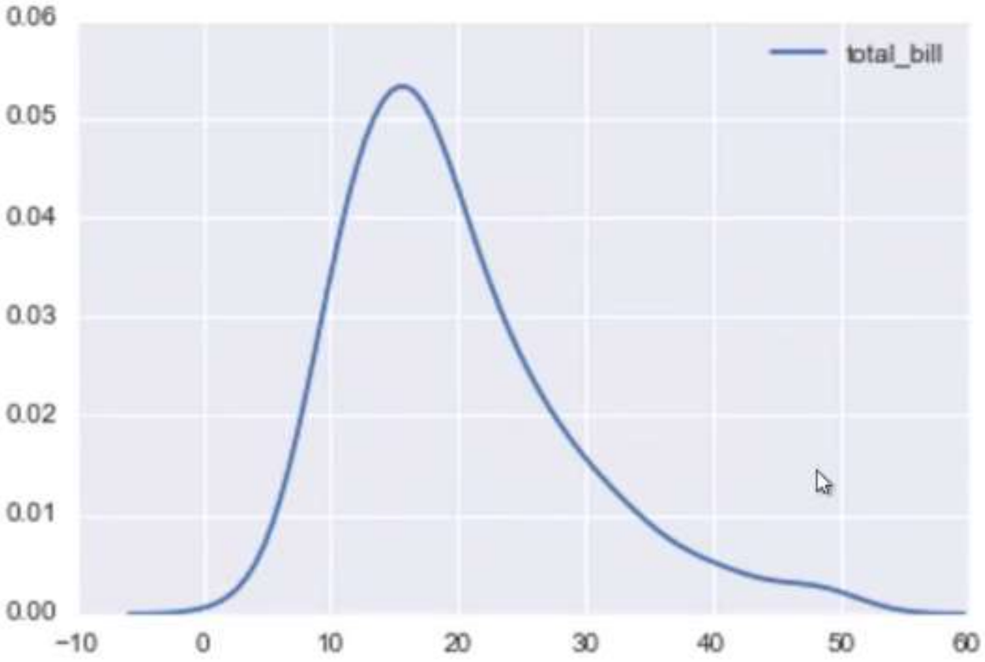
File Edit View Insert Cell Kernel Widgets Help

Python [conda env:py35]

−4 −2 0 2 4 6

In [35]:
```
sns.kdeplot(tips['total_bill'])
```

Out[35]: <matplotlib.axes._subplots.AxesSubplot at 0x254c3635908>



In [ ]: