Seaborn - Visualizing Data Distributions

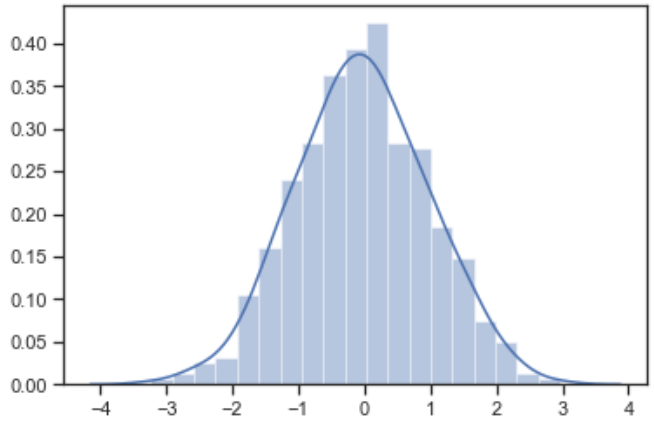
In [1]:
```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

sns.set(style="ticks")
```

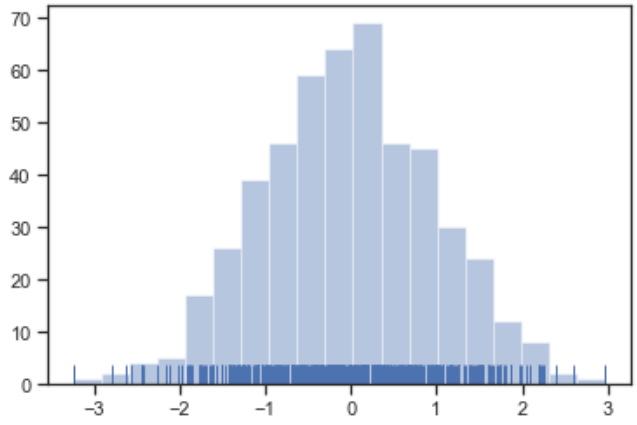
Univariate Distributions

- distplot() - draw a histogram and file a KDE (kernel density estimate)

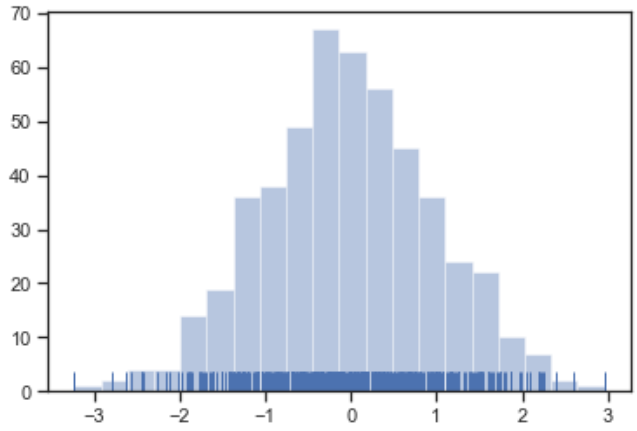
In [3]:
```
np.random.seed(123)
x = np.random.normal(size=500)
sns.distplot(x);
```



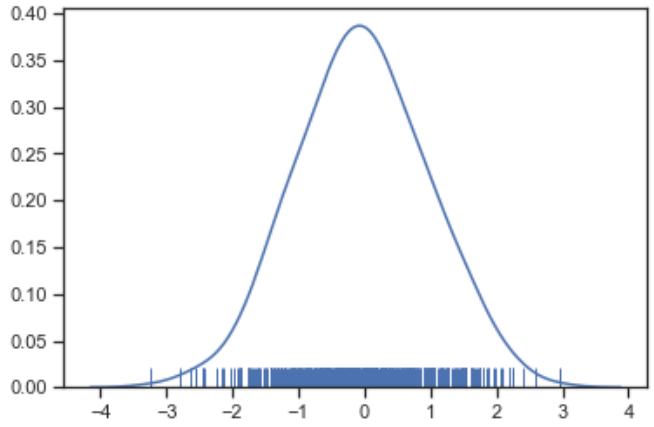
In [4]:
```
sns.distplot(x, kde=False, rug=True);
```



In [5]:
```
sns.distplot(x, bins=20, kde=False, rug=True);
```

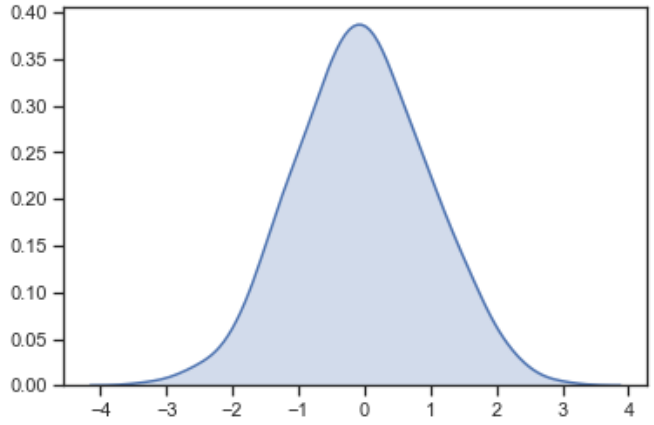


In [6]:
```
sns.distplot(x, bins=20, hist=False, rug=True);
```

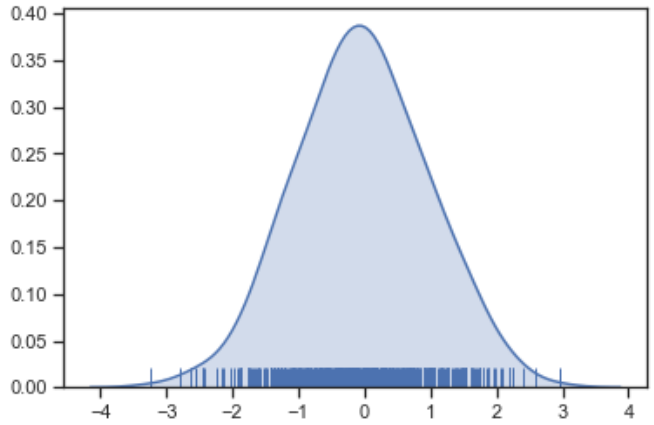


In [7]:
```
# kdeplot provides more options

sns.kdeplot(x, shade=True);
```



In [8]:
```
sns.kdeplot(x, shade=True)
sns.rugplot(x);
```



Bivariate distributions

- jointplot()

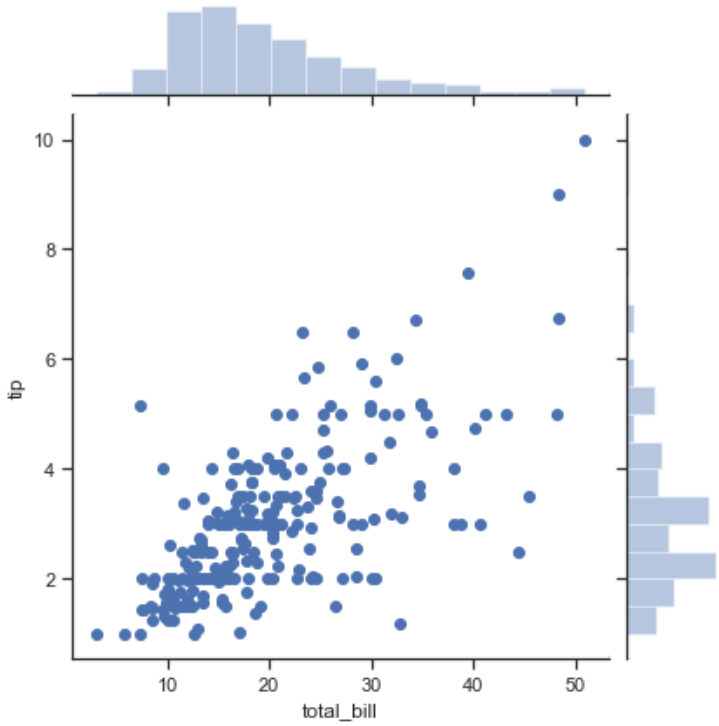
In [9]:
```
tips = sns.load_dataset("tips")
```

In [10]:
```
tips.head()
```

Out[10]:

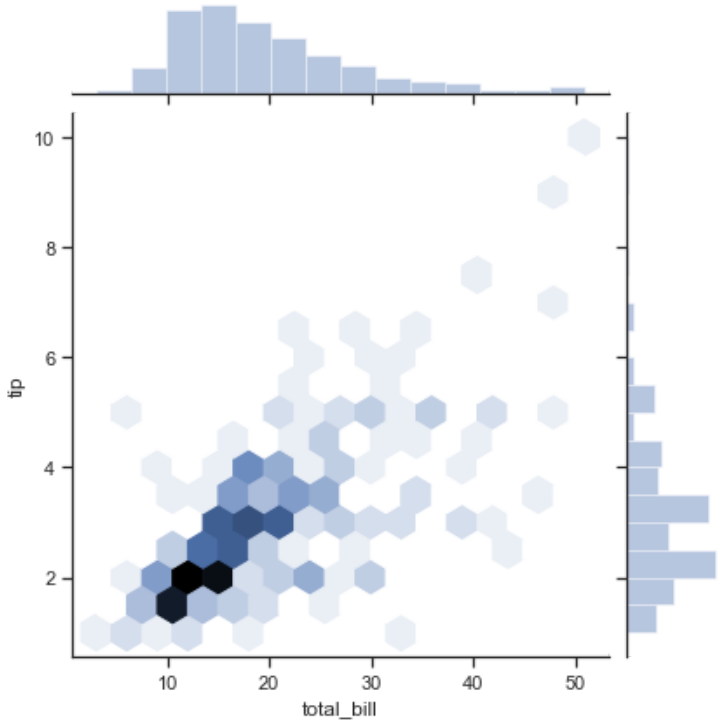
	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4

In [11]:
```
sns.jointplot(x="total_bill", y="tip", data=tips);
```



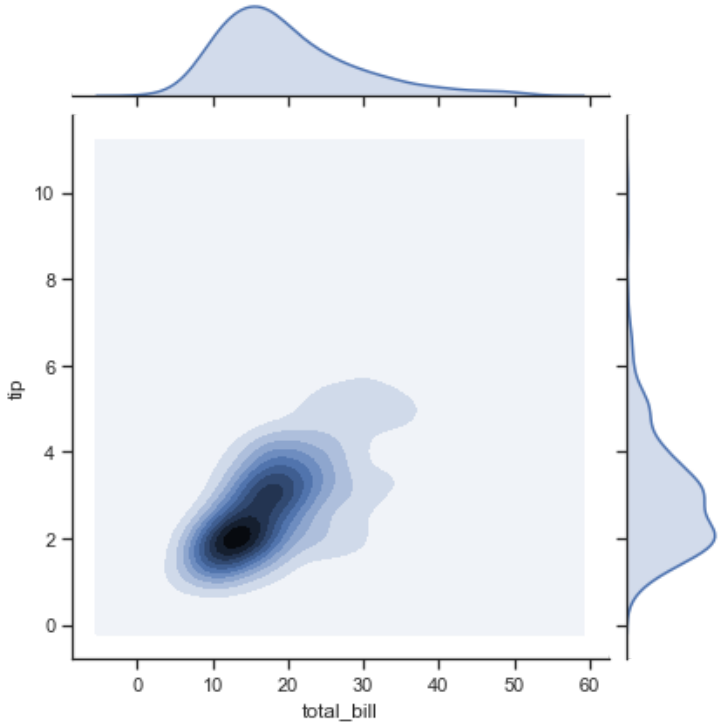
In [12]:
```
# For large datasets, hexbin plot shows counts of observations

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



In [13]:
```
# Use kde to show contour plot

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



Pairwise relationships

In [14]:
```
iris = sns.load_dataset("iris")
iris.head()
```

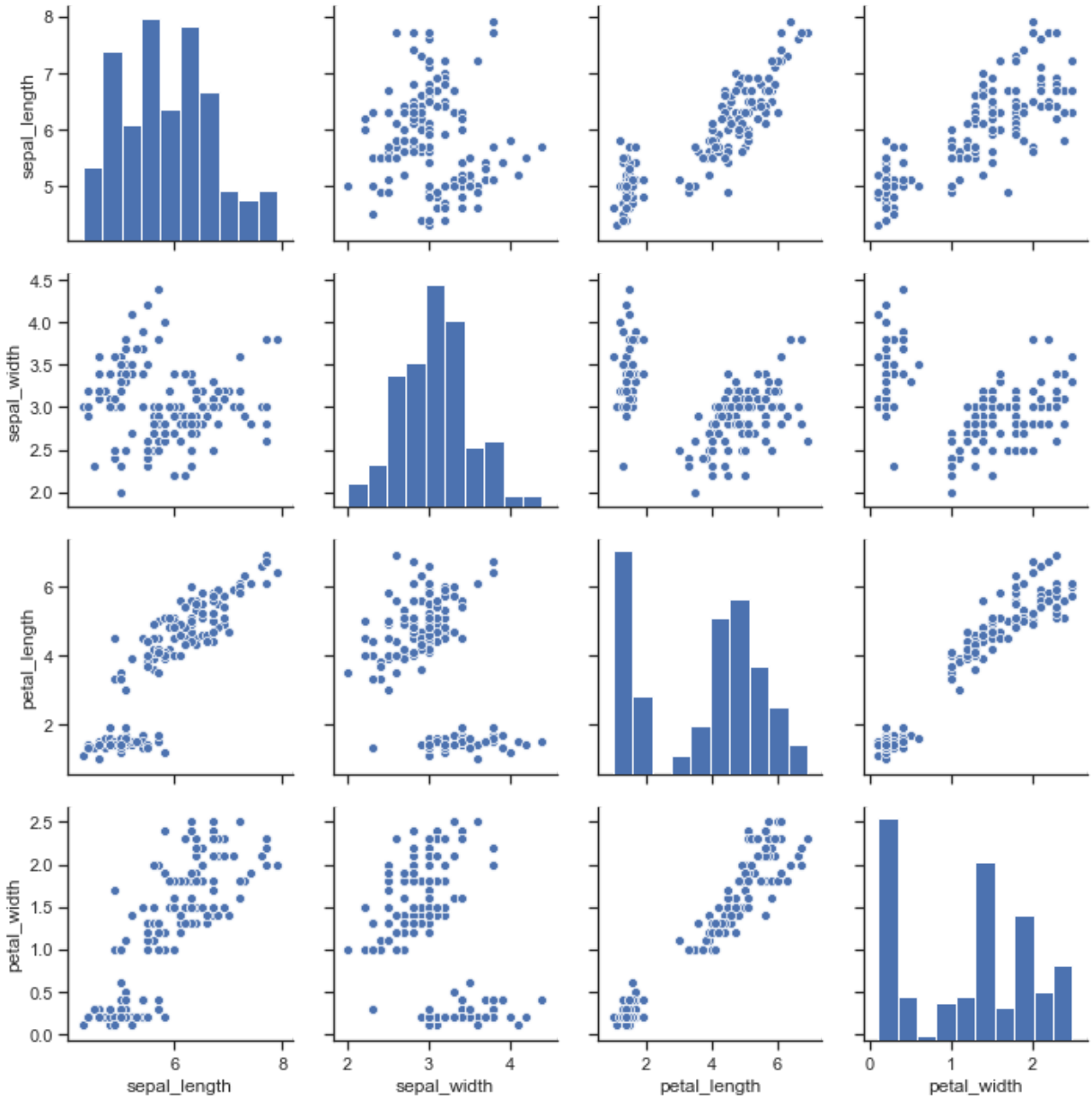
Out[14]:

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa

In [15]:
```
iris.species.unique()
```

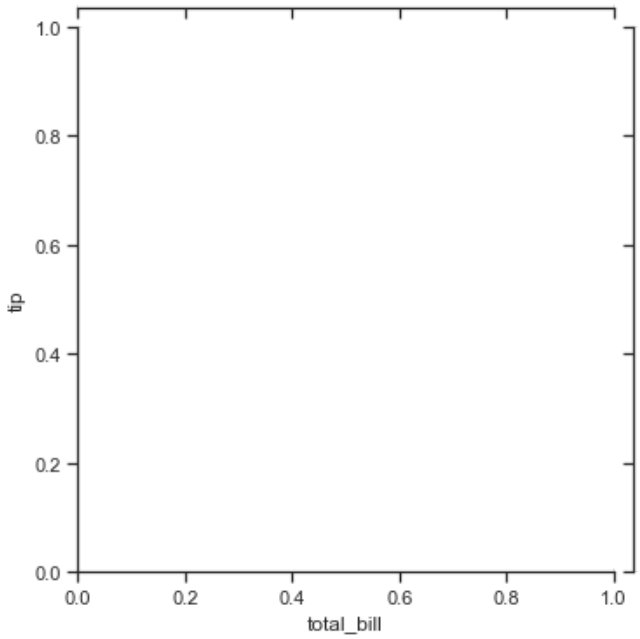
Out[15]:
```
array(['setosa', 'versicolor', 'virginica'], dtype=object)
```

In [16]:
```
sns.pairplot(iris);
```

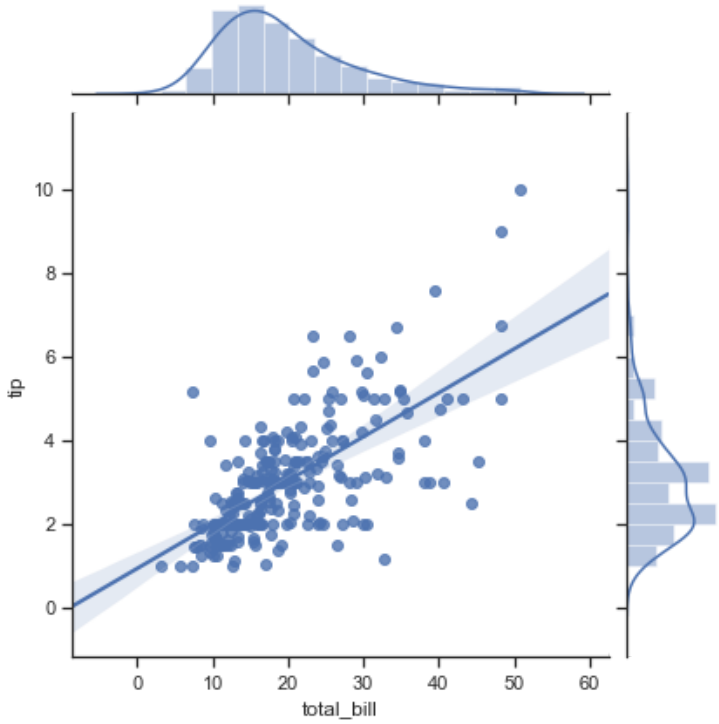


JointGrid

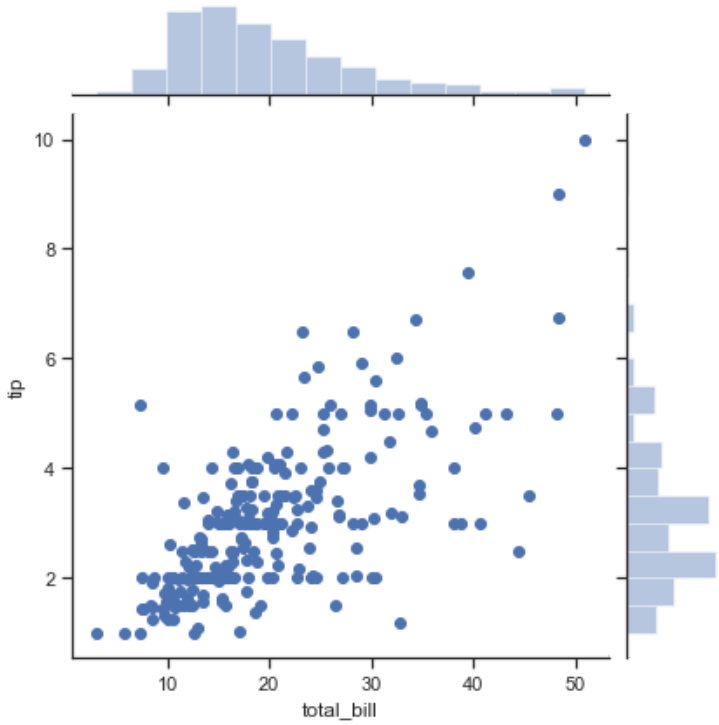
In [17]:
```
g = sns.JointGrid(x="total_bill", y="tip", data=tips)
```



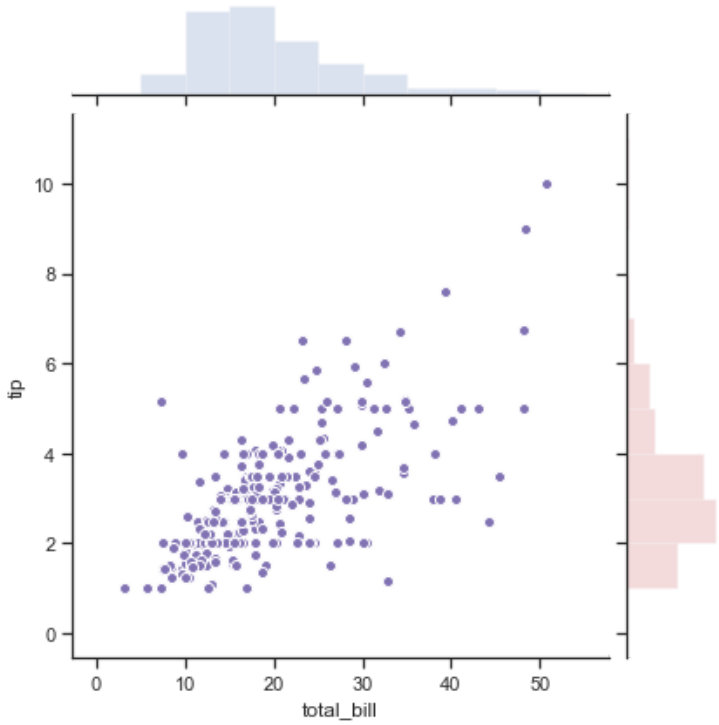
In [18]:
```
g = sns.JointGrid(x="total_bill", y="tip", data=tips)
g = g.plot(sns.regplot, sns.distplot)
```



In [19]:
```
g = sns.JointGrid(x="total_bill", y="tip", data=tips)
g = g.plot_joint(plt.scatter)
g.plot_marginals(sns.distplot, kde=False);
```



In [20]:
```
g = sns.JointGrid(x="total_bill", y="tip", data=tips)
g = g.plot_joint(plt.scatter, color="m", edgecolor="white")
g.ax_marg_x.hist(tips["total_bill"], color="b", alpha=.2,
                 bins=np.arange(0, 60, 5))
g.ax_marg_y.hist(tips["tip"], color="r", alpha=.2,
                 orientation="horizontal",
                 bins=np.arange(0, 12, 1));
```



In [ ]: