Distribution Plots

Data Science Developer

Purwadhika
Startup and Coding School

Distribution Plots

Let's discuss some plots that allow us to visualize the distribution of a data set. These plots are:

- distplot
- jointplot
- pairplot
- rugplot
- kdeplot

Purwadhika
Startup and Coding School

Imports

```python
import seaborn as sns
%matplotlib inline
```

Data

Seaborn comes with built-in data sets!

```python
tips = sns.load_dataset('tips')
```

```python
tips.head()
```

|	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

Purwadhika
Startup and Coding School

distplot

The distplot shows the distribution of a univariate set of observations.

```
sns.distplot(tips['total_bill'])
# Safe to ignore warnings
```

```
C:\Users\harto\Anaconda3\lib\site-packages\matplotlib\axe
s been replaced by the 'density' kwarg.
  warnings.warn("The 'normed' kwarg is deprecated, and ha
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x1e3122b2780>
```
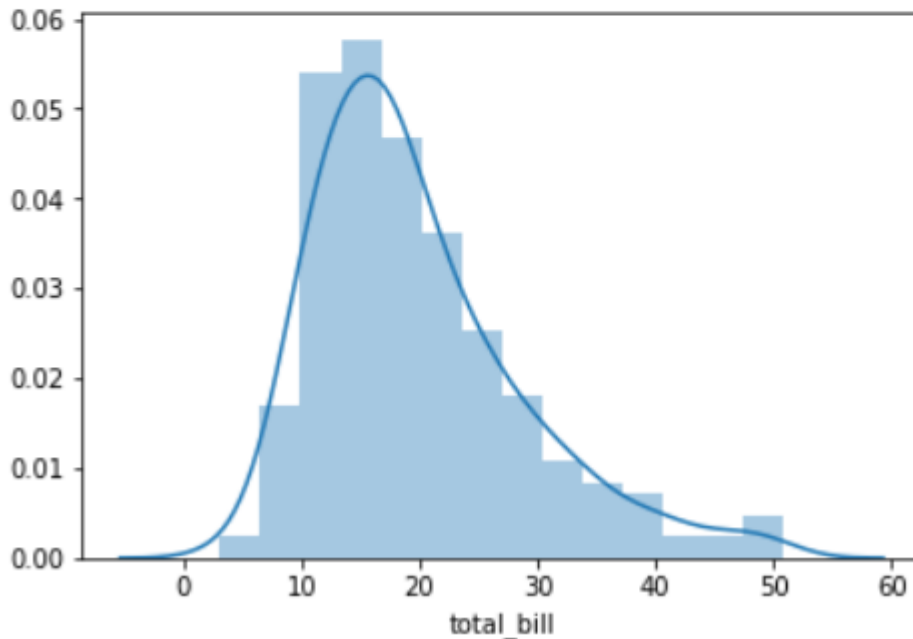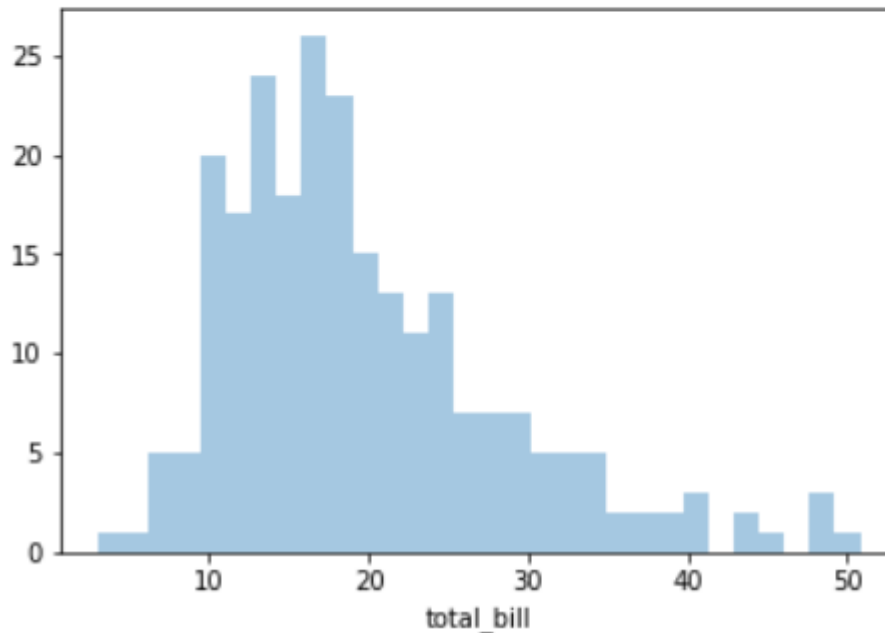
To remove the kde layer and just have the histogram use:

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

```
C:\Users\harto\Anaconda3\lib\site-packages\matplotlib\axe
s been replaced by the 'density' kwarg.
  warnings.warn("The 'normed' kwarg is deprecated, and ha

<matplotlib.axes._subplots.AxesSubplot at 0x1e3124adda0>
```

jointplot

jointplot() allows you to basically match up two distplots for bivariate data. With your choice of what kind parameter to compare with:
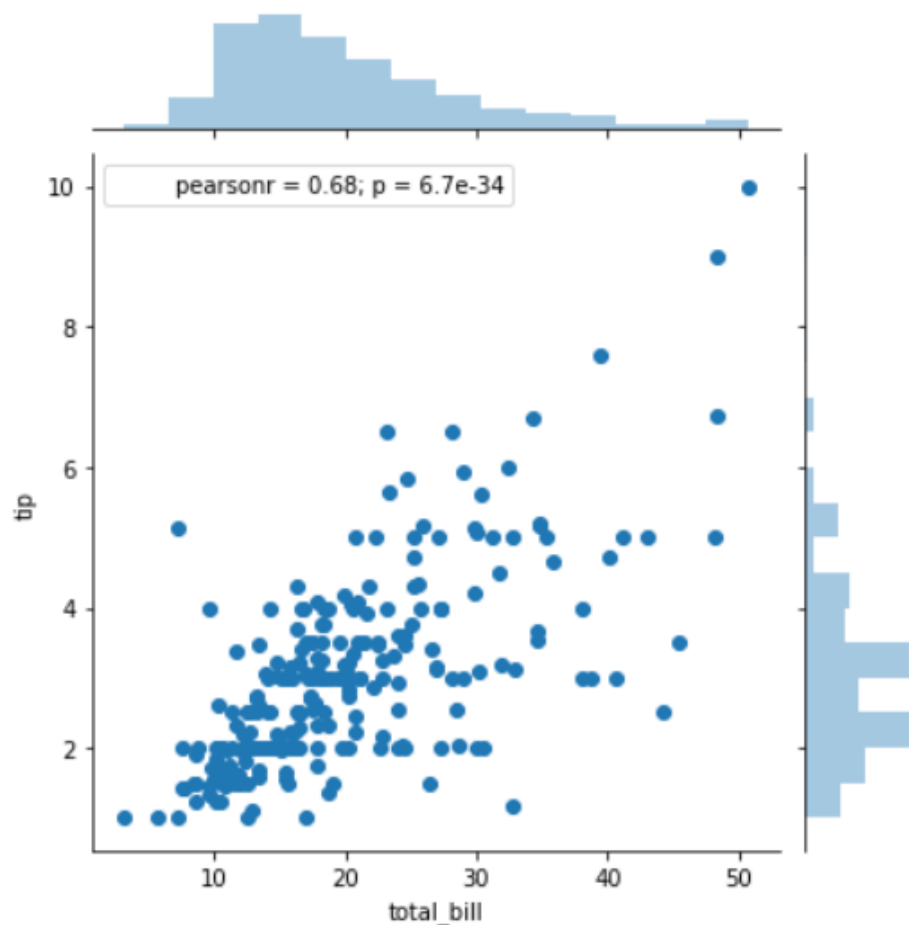
- "scatter"
- "reg"
- "resid"
- "kde"
- "hex"

```
sns.jointplot(x='total_bill',y='tip',data=tips,kind='scatter')
```

C:\Users\harto\Anaconda3\lib\site-packages\matplotlib\axes_axe
s been replaced by the 'density' kwarg.
  warnings.warn("The 'normed' kwarg is deprecated, and has been
C:\Users\harto\Anaconda3\lib\site-packages\matplotlib\axes_axe
s been replaced by the 'density' kwarg.
  warnings.warn("The 'normed' kwarg is deprecated, and has been

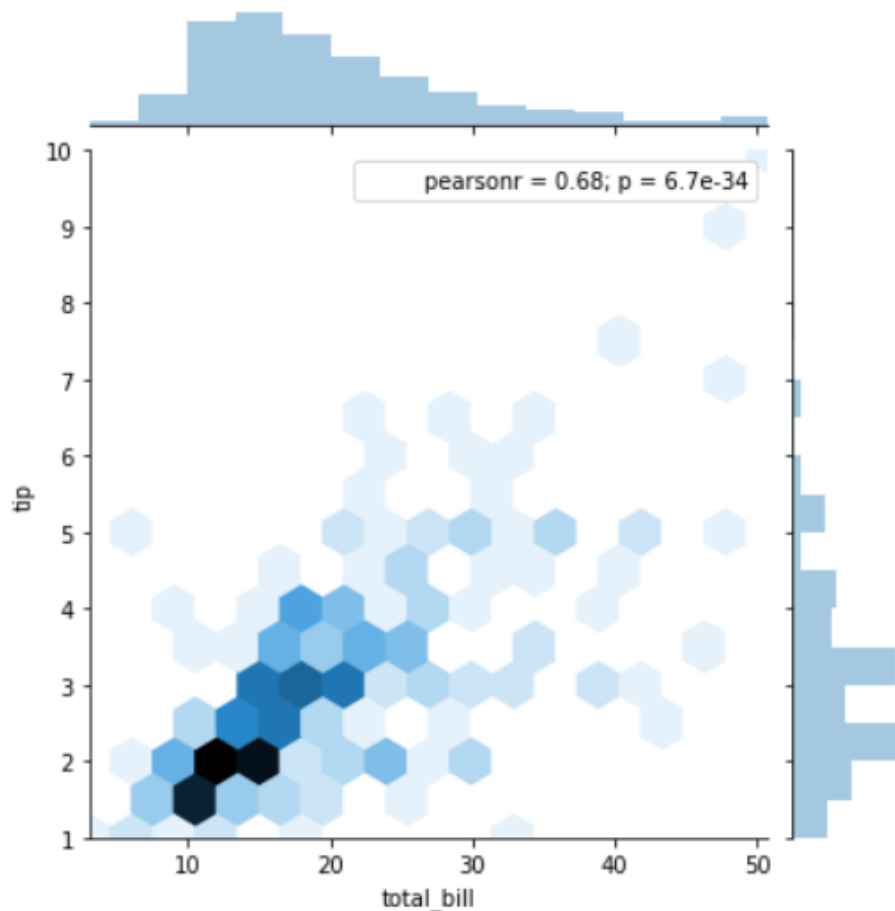<seaborn.axisgrid.JointGrid at 0x1e3127429e8>

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

C:\Users\harto\Anaconda3\lib\site-packages\matplotlib\axes\
s been replaced by the 'density' kwarg.
  warnings.warn("The 'normed' kwarg is deprecated, and has
C:\Users\harto\Anaconda3\lib\site-packages\matplotlib\axes\
s been replaced by the 'density' kwarg.
  warnings.warn("The 'normed' kwarg is deprecated, and has
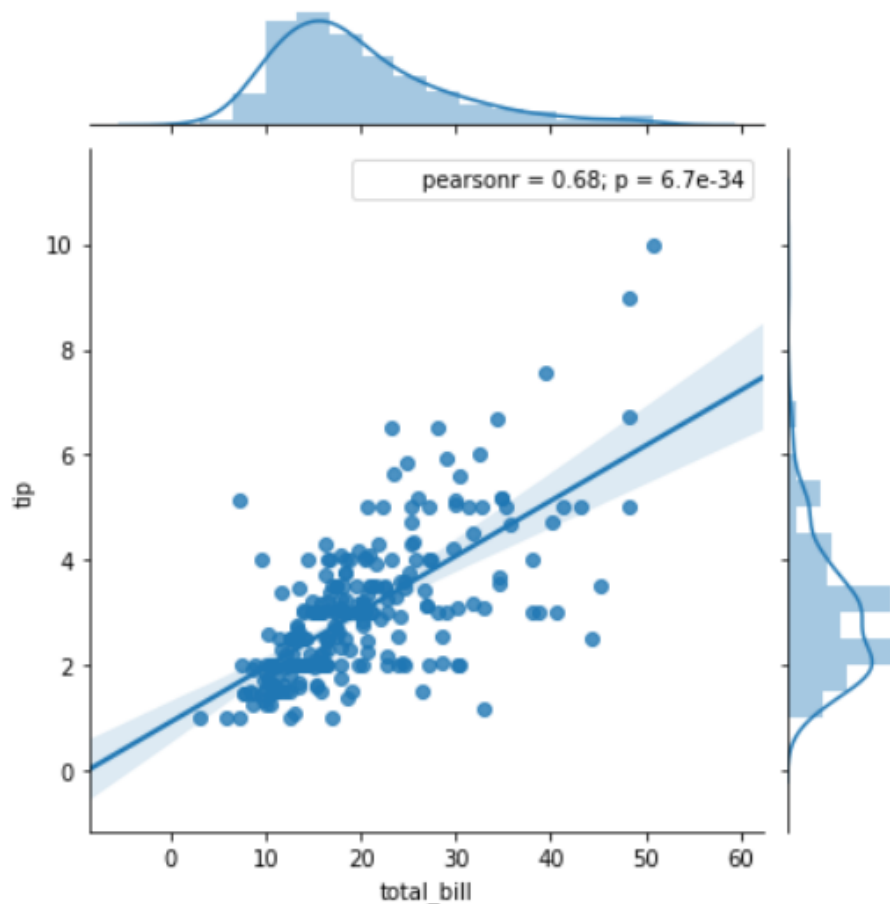
<seaborn.axisgrid.JointGrid at 0x1e313899be0>

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

C:\Users\harto\Anaconda3\lib\site-packages\matplotlib\axes\
s been replaced by the 'density' kwarg.
  warnings.warn("The 'normed' kwarg is deprecated, and has
C:\Users\harto\Anaconda3\lib\site-packages\matplotlib\axes\
s been replaced by the 'density' kwarg.
  warnings.warn("The 'normed' kwarg is deprecated, and has

```
<seaborn.axisgrid.JointGrid at 0x1e3139735c0>
```
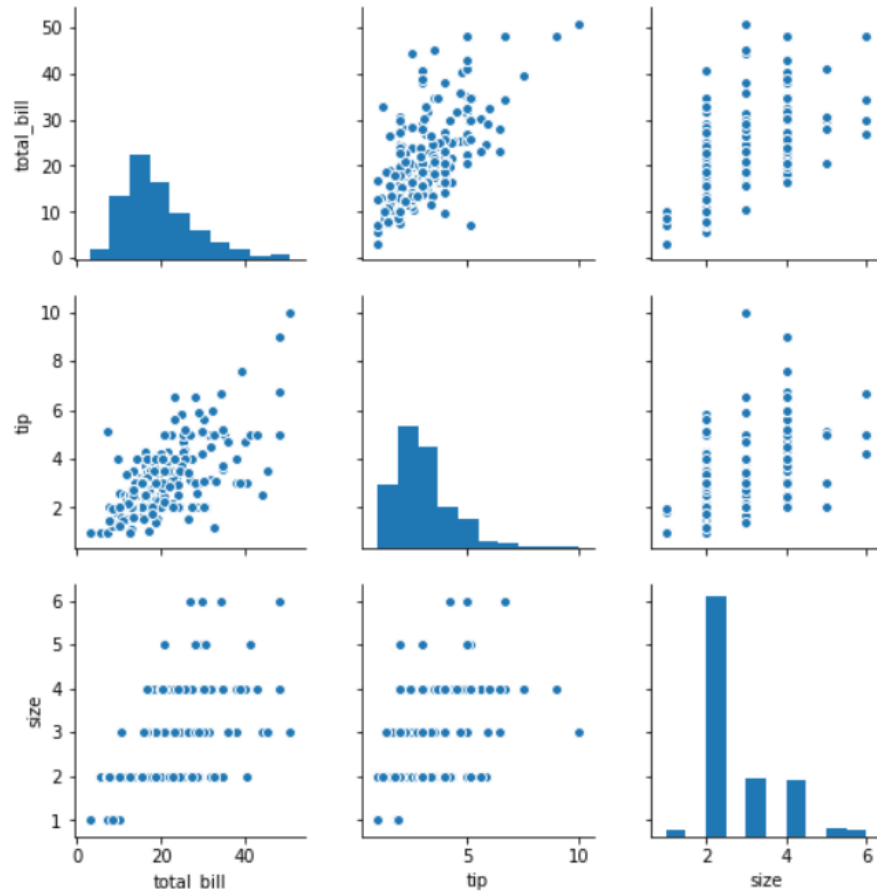
pairplot

pairplot will plot pairwise relationships across an entire dataframe (for the numerical columns) and supports a color hue argument (for categorical columns).
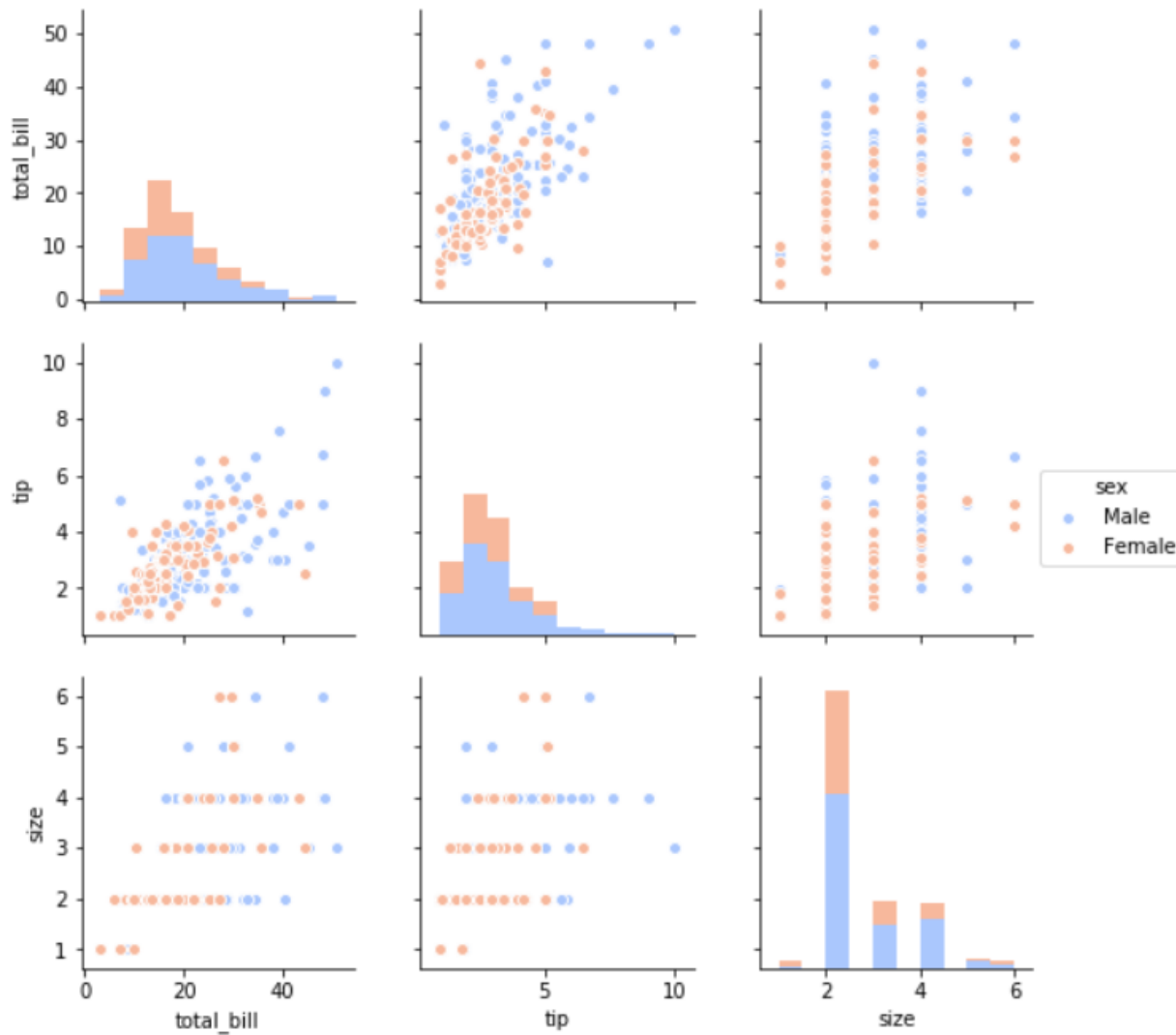
```
sns.pairplot(tips)
```

<seaborn.axisgrid.PairGrid at 0x1e313a9a668>

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

<seaborn.axisgrid.PairGrid at 0x1e313fb2630>

rugplot

rugplots are actually a very simple concept, they just draw a dash mark for every point on a univariate distribution. They are the building block of a KDE plot:

```
sns.rugplot(tips['total_bill'])
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x21aaad374a8>
```


Purwadhika
Startup and Coding School