

Getting Started with Seaborn : Basics Tutorial

Author, Justine Miranda

A Data Science Foundation Blog

May 2020

www.datascience.foundation

Copyright 2016 - 2017 Data Science Foundation

Python contains numerous data visualization tools for you to enhance & create some beautiful graphs, plots & charts. When I look at those visualization tools, one library stands out the most and that is 'Seaborn'. This visualization tool is very powerful in terms of its application, its quite easy to use & it combines some aesthetic appeal to the graphs you design.

In this article, I am going to run through some most simplest & basic graphs that you can plot using Seaborn. This will give you a better picture in order to get hands on experience with this library.

Let's get started with importing our Seaborn library.

```
import seaborn as sns
```

Once you have imported the library, you can set the style for all your plots & graphs. There are quite a few, which I will discuss in detail. But you can use the below to get started with -

```
sns.set_style('whitegrid')
```

Alright! So now, your library has been imported & you have set your seaborn background style. Now, we need to import data to work with, in order to create some extraordinary simple plots.

Note—Seaborn comes with some in-built data sets that you can work with.

Here, we are importing 'tips' dataset to give you an idea on how we are going to work further. Th data is all about people having food at a restaurant & whether or not they left a tip. Importing dataset is quite simple with a one-liner code that goes as follow —

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

I am setting my dataframe as 'tips' here.

Great! So, we are good to proceed ahead showing you few examples of seaborn visualization. I have categorized seaborn into four plots -

1. **Distribution Plots**
2. **Categorical Plots**
3. **Matrix Plots**
4. **Regression Plots**

Further to this four plots, there are plenty others. But, in this article, I am going to focus on the types of 'Distribution Plots' & is often simply to plot with one-liner codes.

Let's get started with our first type -

Distribution Plots

Distribution Plots are nothing but a distribution of a sample data plotted by comparing the empirical distribution of the data with numerical/theoretical values. It is widely used in order to compare range and distribution.

Let's have a look at all the sub-types under this category —

DistPlot

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

Let's have a look at our data first before plotting Distplot.

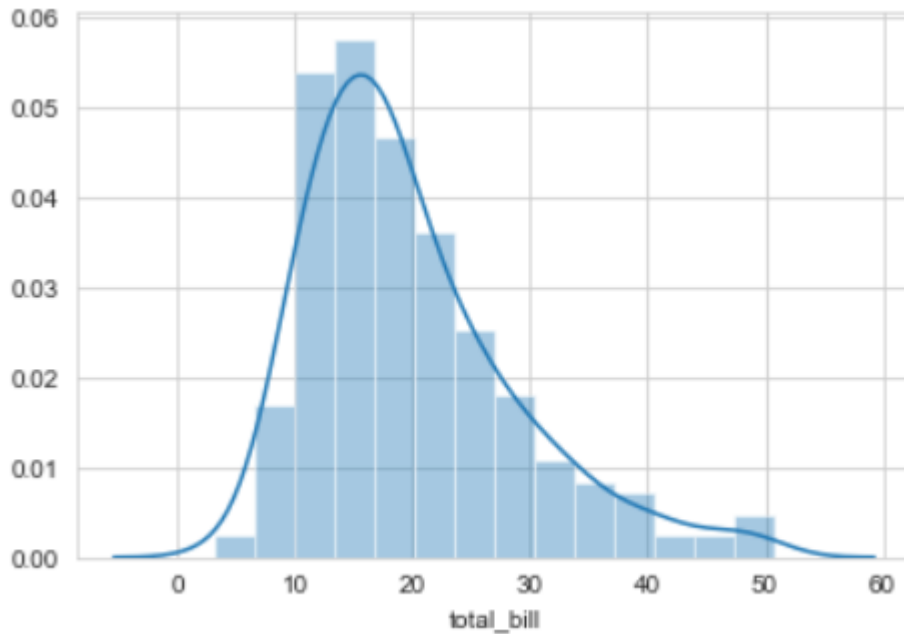
```
tips.head()
```

```
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

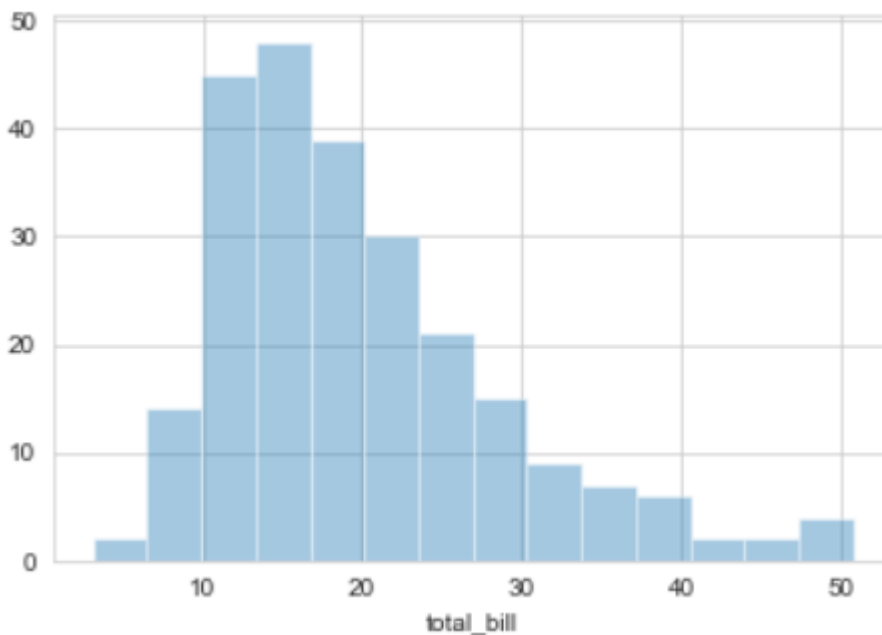
As shown in the above data, we have various categories & groups stating the gender, the time of the day & tips drawn with the total bill. Let's use this data to plot our Distplot.

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



As you can see, we have plotted our Distplot which shows the distribution & its range in a set of observations. The curve plotted over the distribution is called 'kde'. You can completely remove the curve by passing -

```
sns.distplot(df['total_bill'], kde = False)
```



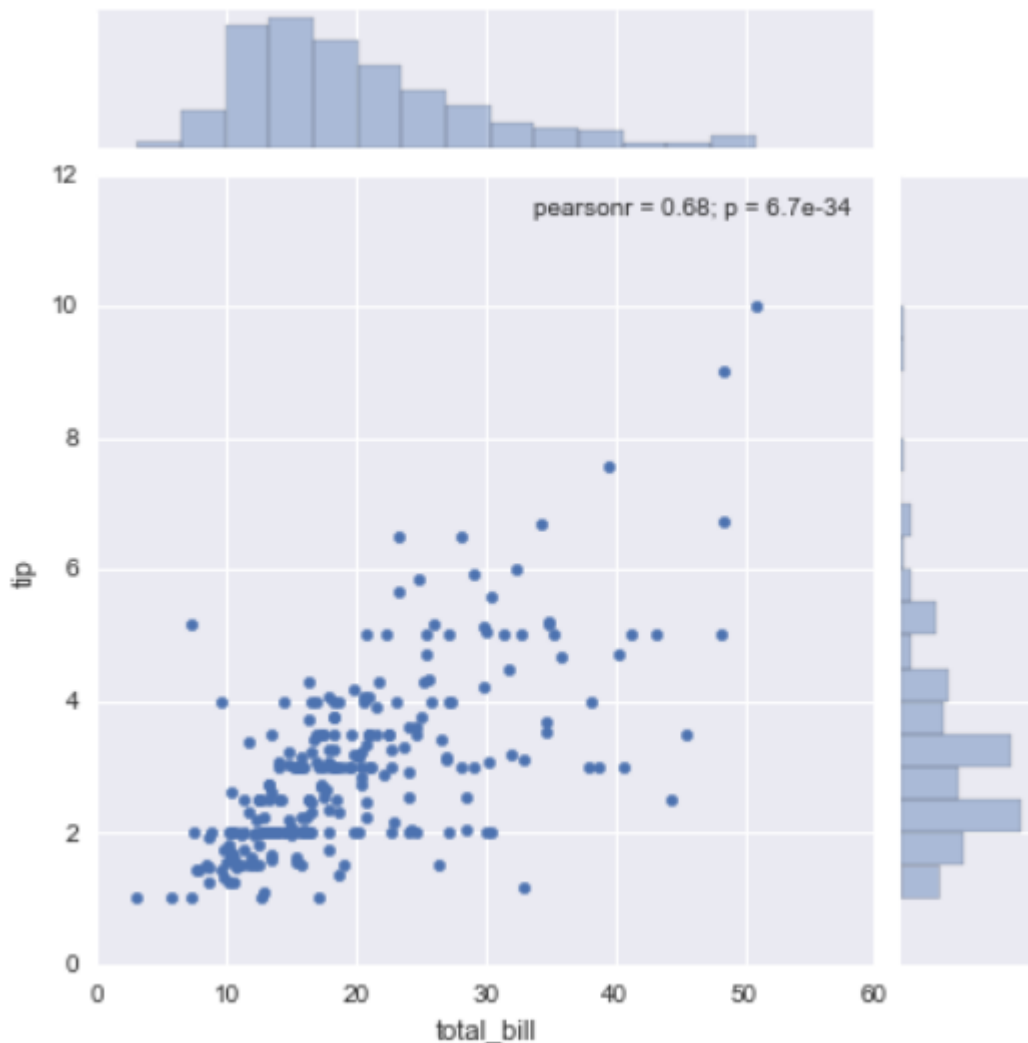
And now, the kde curve has been removed.

Joint Plot

Joint Plot allows you to basically match up two distplots for bivariate data.

In Joint Plot, you have to pass the values of both 'x' & 'y'.

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



As you can see, by default scatter plot has been plotted. You can change it & use various kinds to customize your plot. The following values can be used in order to visualize —

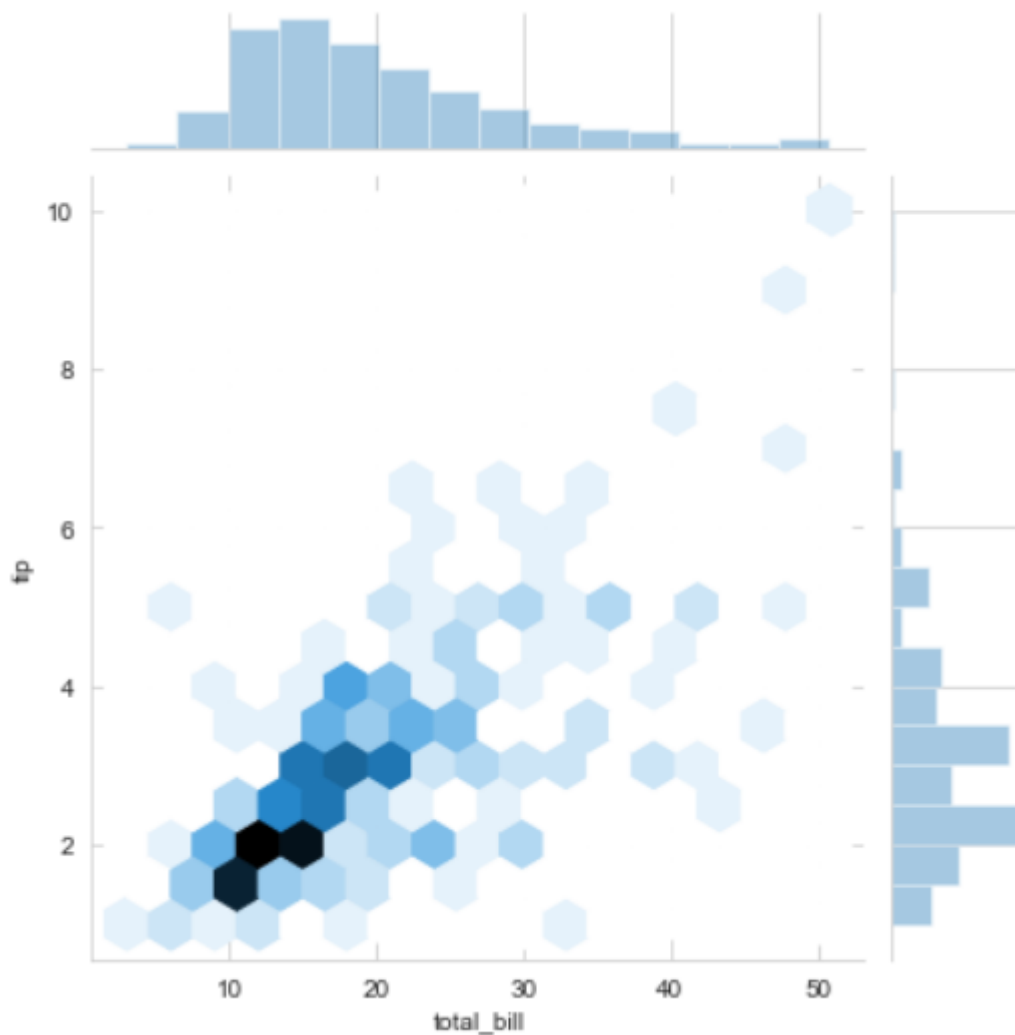
- scatter

Data Science Foundation

- reg
- resid
- kde
- hex

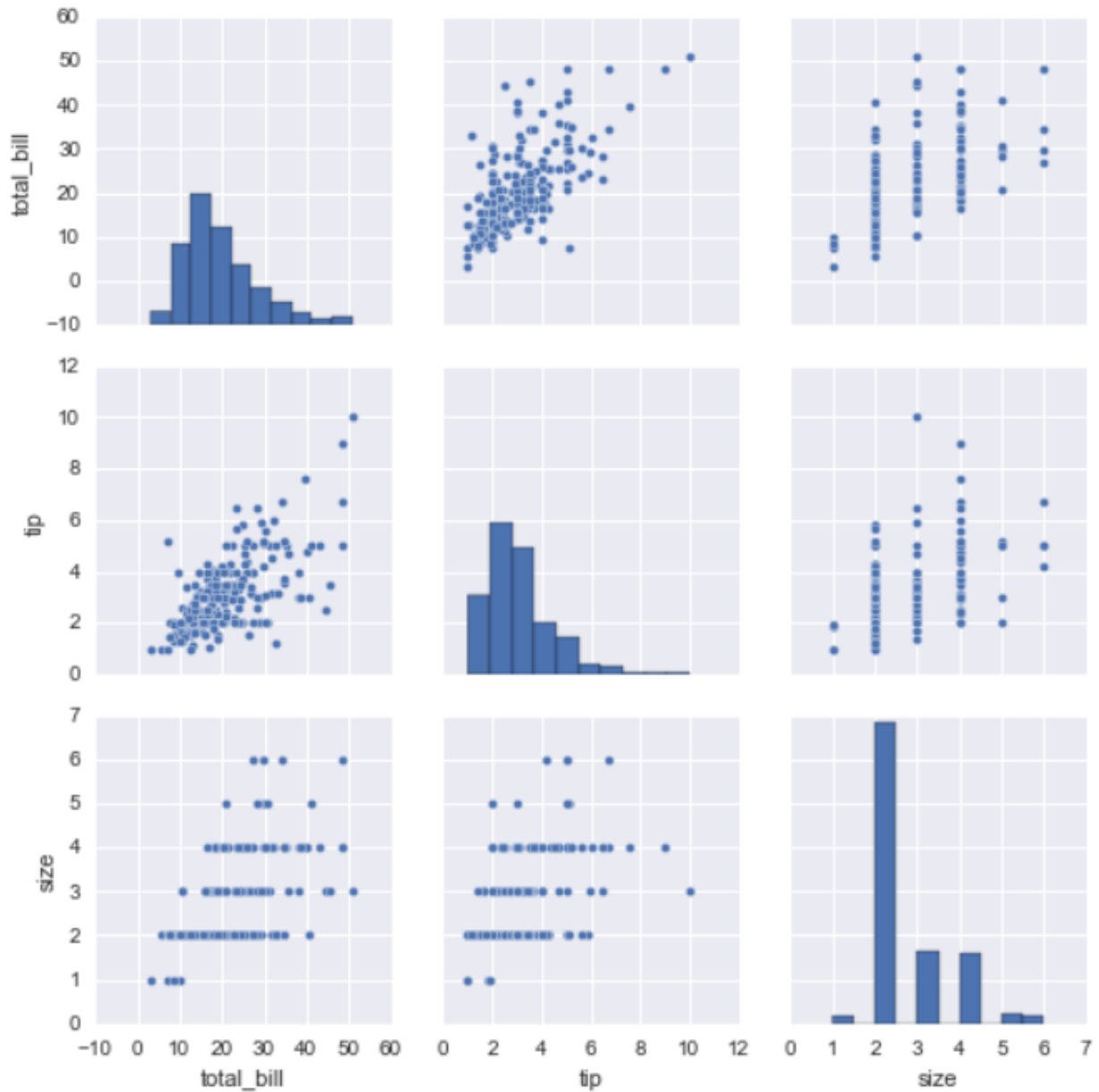
Let me show you one more example on how to plot such distribution levels by passing in the values.

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



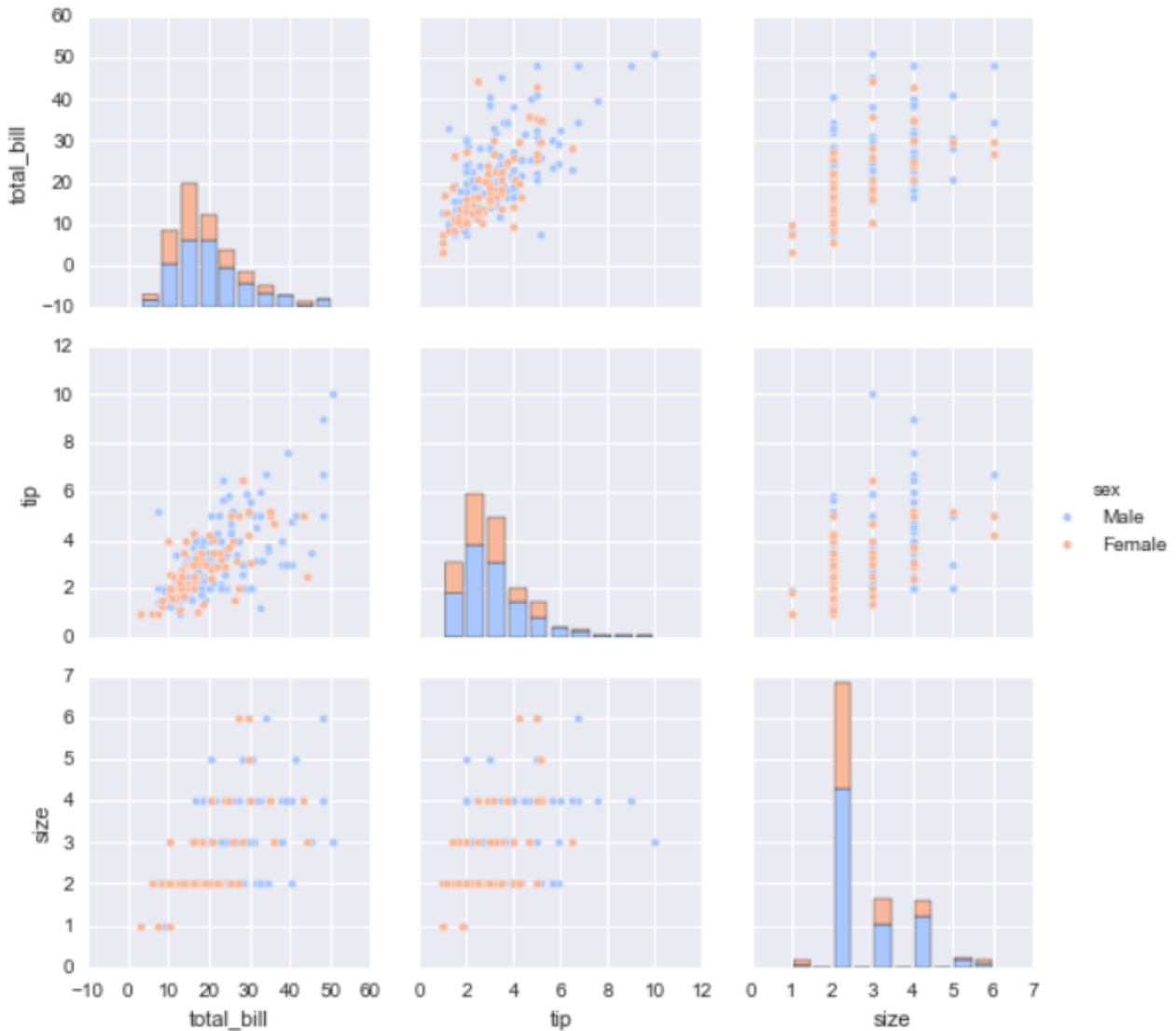
Pairplot will plot pairwise relationships across an entire dataframe (for the numerical columns) and supports a color hue argument (for categorical columns).

```
sns.pairplot(tips)
```



To visualize it better, you can use various colors & style to give an aesthetic look.

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



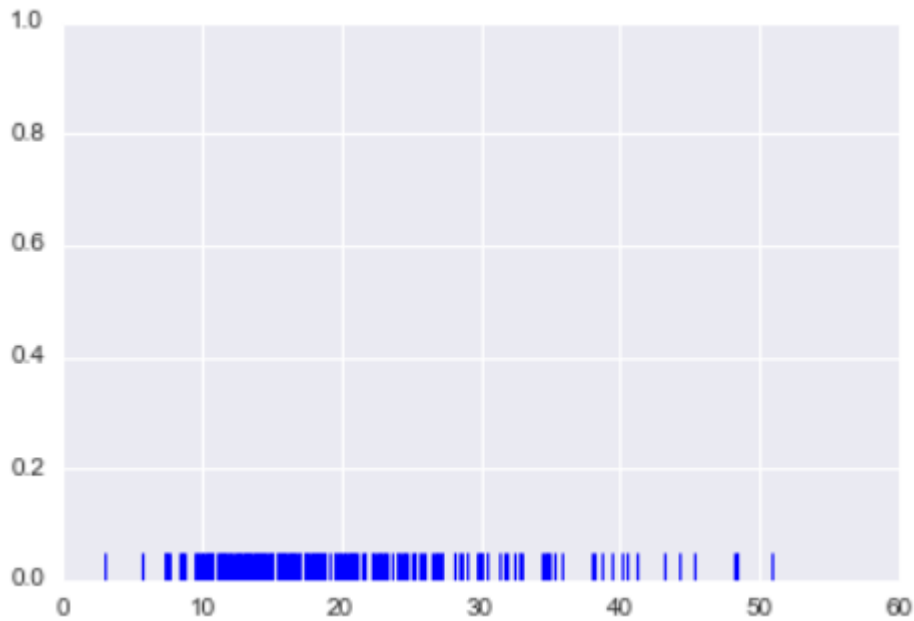
For different sorts of palettes, you can visit the seaborn official document page —

[Choosing color palettes - seaborn 0.10.1 documentation](#)

[Color is more important than other aspects of figure style because color can reveal patterns in the data if used...seaborn.pydata.org](#)

Rug plot

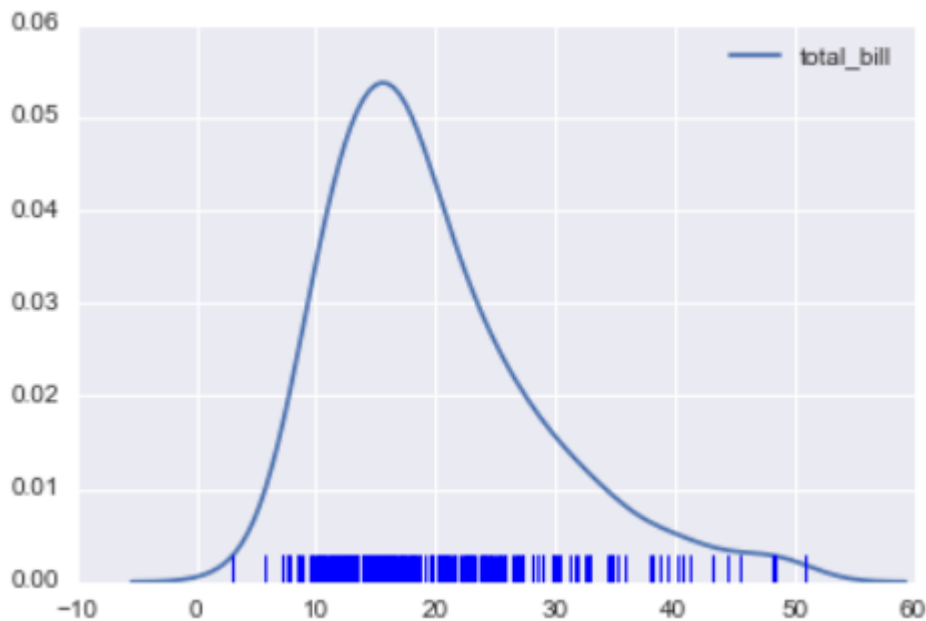
Rugplot are actually a very simple concept, they just draw a dash mark for every point on a univariate distribution.



KDE plot

kde plots are known as 'Kernel Density Estimation Plots. These KDE plots replace every single observation with a Gaussian (Normal) distribution centered around that value.

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



Great! I hope this article helped you all in understanding the basic plotting scales of 'Distribution' type. It

is widely used & very simple to understand by any user who wants to grab some real quick information present in data.

Thanks for taking out time & reading this article!

Hope you enjoyed!

About the Data Science Foundation

The Data Science Foundation is a professional body representing the interests of the Data Science Industry. Its membership consists of suppliers who offer a range of big data analytical and technical services and companies and individuals with an interest in the commercial advantages that can be gained from big data. The organisation aims to raise the profile of this developing industry, to educate people about the benefits of knowledge based decision making and to encourage firms to start using big data techniques.

Contact Data Science Foundation

Email: admin@datascience.foundation
Telephone: 0161 926 3641
Atlantic Business Centre
Atlantic Street
Altrincham
WA14 5NQ
web: www.datascience.foundation

Data Science Foundation

Data Science Foundation, Atlantic Business Centre, Atlantic Street, Altrincham, WA14 5NQ
Tel: 0161 926 3670 Email: admin@datascience.foundation Web: www.datascience.foundation
Registered in England and Wales 4th June 2015, Registered Number 9624670