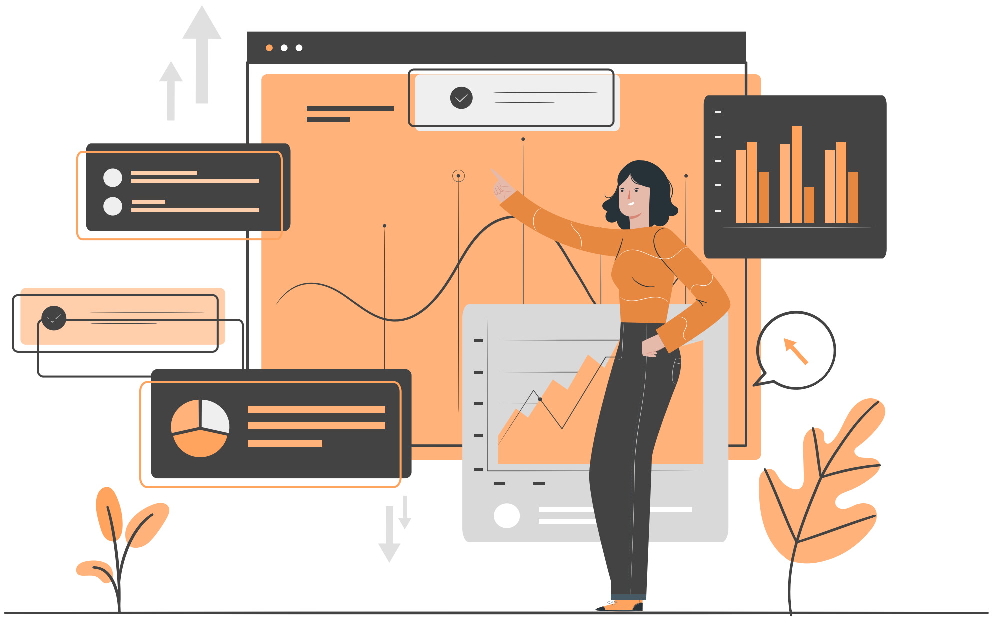


Seaborn

By Fireblaze AI School

Index

Index

Visualization

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.

Introduction

- In the world of Analytics , the best way to get insights is by visualizing the data. Data can be visualized by representing it as plots which is easy to understand, explore and grasp. Such data helps in drawing the attention of key elements.
- To analyse a set of data using Python, we make use of Matplotlib, a widely implemented 2D plotting library. Likewise, Seaborn is a visualization library in Python. I t is built on top of Matplotlib.

Matplotlib vs Seaborn

I t is summarized that if Matplotlib “tries to make easy things easy and hard things possible”, Seaborn tries to make a well-defined set of hard things easy too.” Seaborn helps resolve the two major problems faced by Matplotlib; the problems are:

- Default Matplotlib parameters
- Working with data frames

Matplotlib Vs Seaborn

As Seaborn compliments and extends Matplotlib, the learning curve is quite gradual. If you know Matplotlib, you are already half way through Seaborn.

Important Features of Seaborn

- Built in themes for styling matplotlib graphics
- Visualizing univariate and bivariate data
- Fitting in and visualizing linear regression models
- Plotting statistical time series data
- Seaborn works well with NumPy and Pandas data structures
- It comes with built in themes for styling Matplotlib graphics

Installing Seaborn & getting Started

Installing Seaborn and getting started

In this section, we will understand the steps involved in the installation of Seaborn.

Using Pip Installer

To install the latest release of Seaborn, you can use pip:

```
pip install seaborn
```

View Seaborn Dataset

To view all the available data sets in the Seaborn library, you can use the following command with the **get_dataset_names()** function as shown below:

```
import seaborn as sb

 print (sb.get_dataset_names())
```

The above line of code will return the list of datasets available as the following output **[u'anscombe', u'attention', u'brain_networks', u'car_crashes', u'dots', u'exercise', u'flights', u'fmri', u'gammas', u'iris', u'planets', u'tips', u'titanic']**

Displot

What it is?

This function provides access to several approaches for visualizing the univariate or bivariate distribution of data, including subsets of data defined by semantic mapping and faceting across multiple subplots.

displot

What Does It Visualize?

A Displot or distribution plot, **depicts the variation in the data distribution**. Seaborn Distplot represents the overall distribution of continuous data variables. The Seaborn module along with the Matplotlib module is used to depict the distplot with different variations in it.

displot( )

What Does It Measure?

Seaborn distplot lets you show **a histogram with a line on it**. This can be shown in all kinds of variations. We use seaborn in combination with matplotlib, the Python plotting module. A distplot plots a univariate distribution of observations.

displot( )

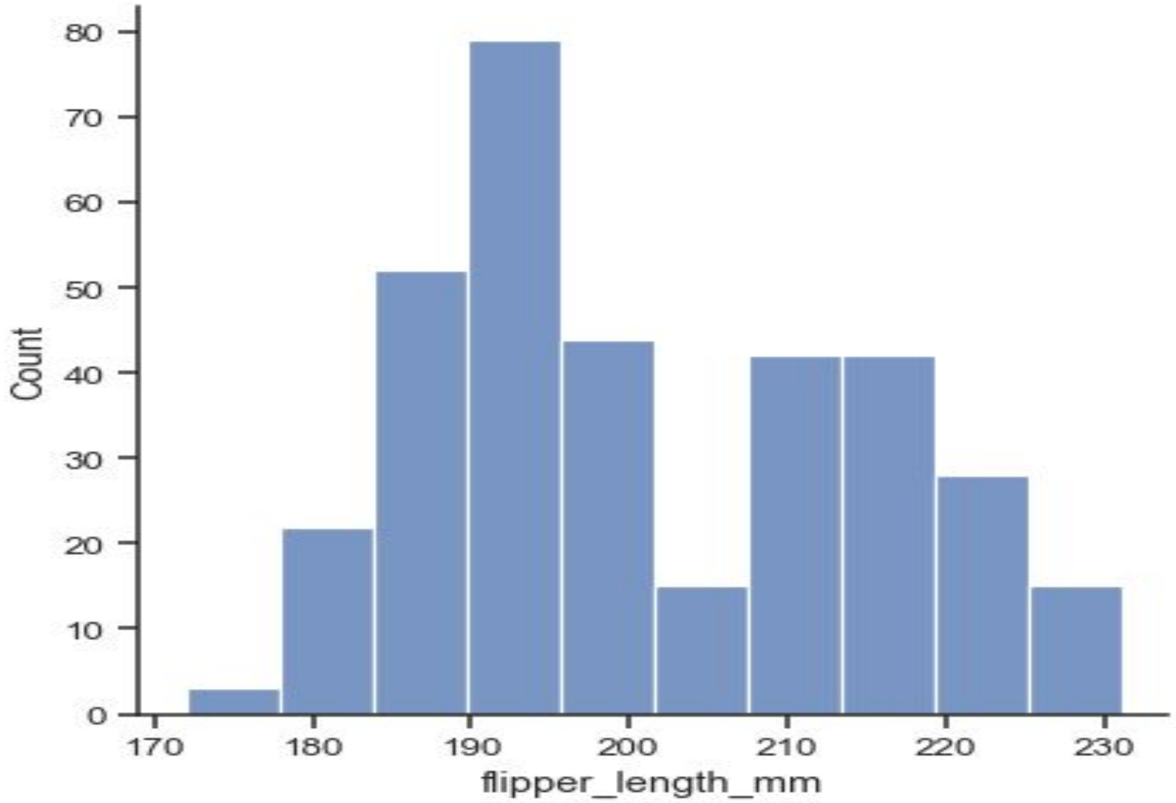
Examples

See the API documentation for the axes-level functions for more details about the breadth of options available for each plot kind.

The default plot kind is a histogram:

```
penguins = sns.load_dataset("penguins")

sns.displot(data=penguins, x="flipper_length_mm")
```

displot( )

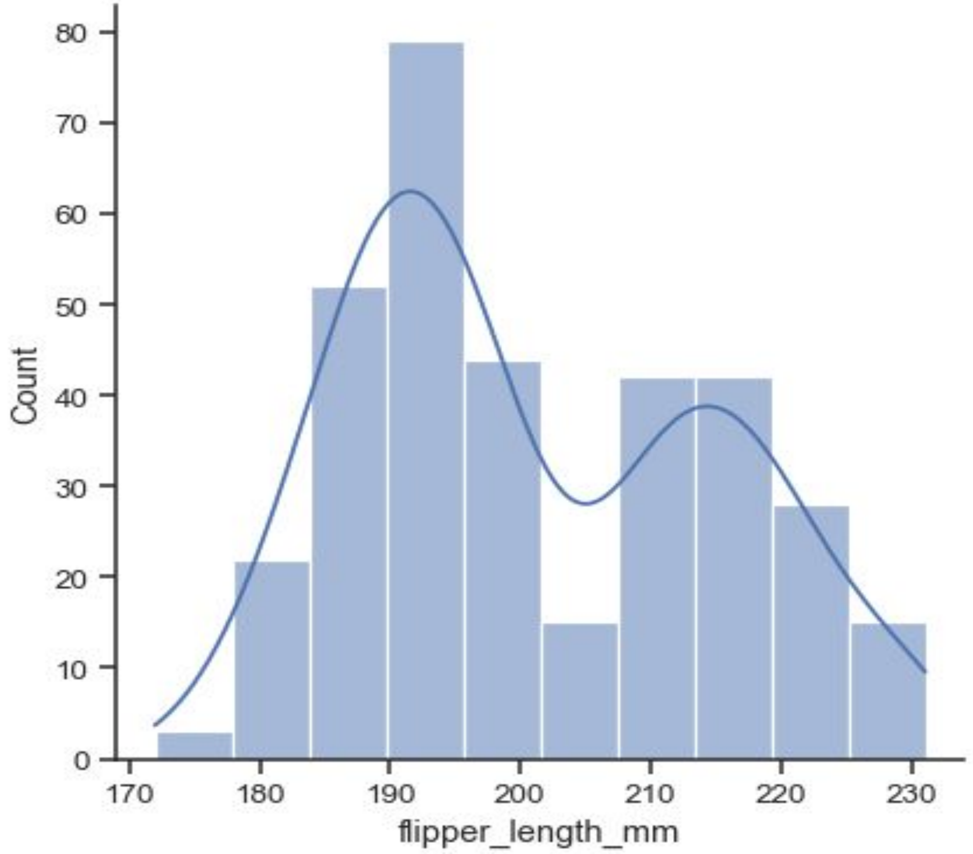


Count
80
70
60
50
40
30
20
10
0
170
180
190
200
210
220
230
flipper_length_mm

While in histogram mode, it is also possible to add a KDE curve:

```
sns.displot(data=penguins, x="flipper_length_mm", kde=True)
```

displot( )


Count
80
70
60
50
40
30
20
10
0
170
180
190
200
210
220
230
flipper_length_mm

Histogram

What is It?

A histogram is **a bar graph-like representation of data that buckets a range of outcomes into columns along the x-axis**. The y-axis represents the number count or percentage of occurrences in the data for each column and can be used to visualize data distributions.

Histogram

What Does It Visualize?

A Histogram visualises **the distribution of data over a continuous interval or certain time period**. Each bar in a histogram represents the tabulated frequency at each interval/bin.

Histogram

What Does It Measure?

A histogram is a type of graph that has wide applications in statistics. Histograms **provide a visual interpretation of numerical data by indicating the number of data points that lies within a range of values**. These ranges of values are called classes or bins.

Histogram

What Sources of Data Does It Use?

The histogram is a popular graphing tool. It is used to summarize **discrete or continuous data** that are measured on an interval scale. It is often used to illustrate the major features of the distribution of the data in a convenient form

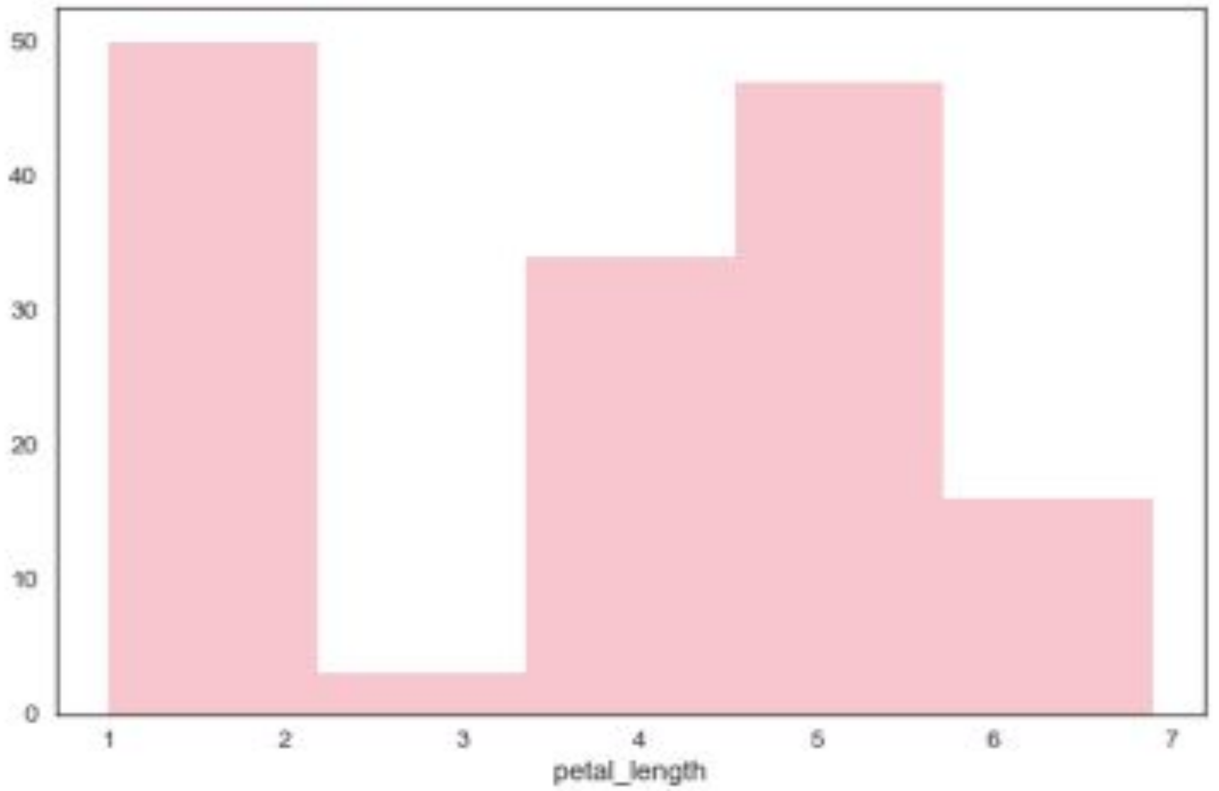
Histogram

Example

```
import pandas as pd

import seaborn as sb

from matplotlib import pyplot as plt

df = sb.load_dataset('iris')

 sb.histplot(df['petal_length'])

 plt.show()
```

Histogram

Output :



KDE

What is It?

A kernel density estimate (KDE) plot is **a method for visualizing the distribution of observations in a dataset, analagous to a histogram**. KDE represents the data using a continuous probability density curve in one or more dimensions.

KDE

What Does It Visualize?

A kernel density estimate (KDE) plot is a method for visualizing **the distribution of observations in a dataset, analagous to a histogram**. KDE represents the data using a continuous probability density curve in one or more dimensions.

KDE

What Does It Measure?

Kernel density estimation is a technique for **estimation of probability density function** that is a must-have enabling the user to better analyse the studied probability distribution than when using a traditional histogram.

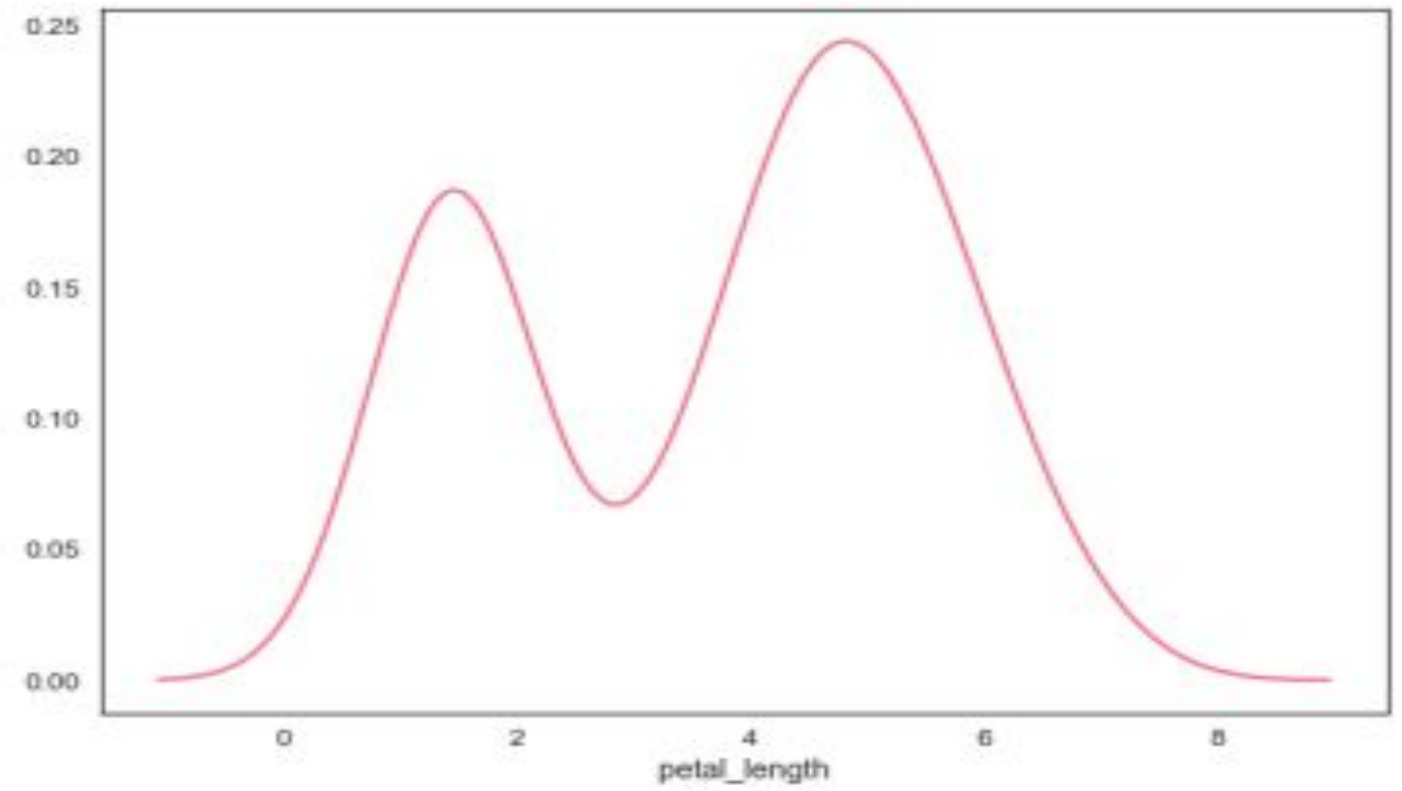
KDE

Example

```
import pandas as pd

import seaborn as sb

from matplotlib import pyplot as plt

df = sb.load_dataset('iris')

sb.kdeplot(df['petal_length'])

plt.show()
```

KDE

Output :



Scatterplot

What is It?

A scatter plot (aka scatter chart, scatter graph) **uses dots to represent values for two different numeric variables**. The position of each dot on the horizontal and vertical axis indicates values for an individual data point. Scatter plots are used to observe relationships between variables.

Scatterplot

What Does It Visualize?

A scatter plot is a type of data visualization that shows **the relationship between different variables**. This data is shown by placing various data points between an x- and y-axis. Essentially, each of these data points looks “scattered” around the graph, giving this type of data visualization its name

Scatterplot

What Does It Measure?

Use a scatter plot to determine **whether or not two variables have a relationship or correlation**. Are you trying to see if your two variables might mean something when put together? Plotting a scattergram with your data points can help you to determine whether there's a potential relationship between them

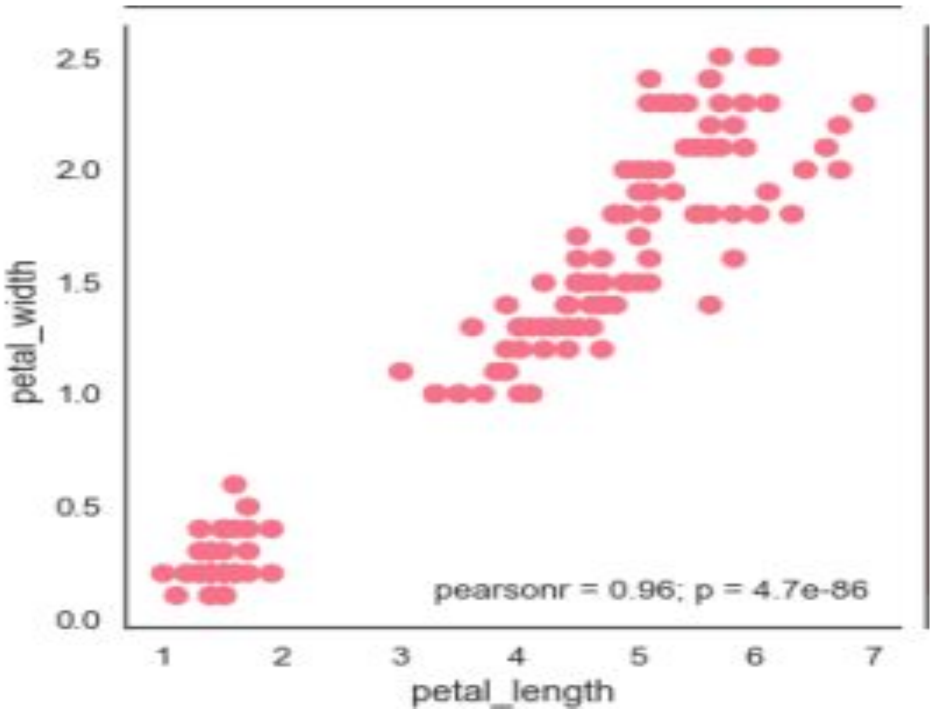
Scatterplot

Example

```
import pandas as pd

import seaborn as sb

from matplotlib import pyplot as plt

df = sb.load_dataset('iris')

sb.scatterplot(x='petal_length',y='petal_width',data=df)

plt.show()
```

Scatterplot

Output:



Barplot

What is It?

A bar chart is essentially a column chart on its side: values are presented on the horizontal axis and the categories are on vertical axis, on the left.

Barplot

What Does It Visualize?

Bar charts are more commonly used to compare different values, items and categories of data. From a purely practical perspective, they're also used over column charts when the names of the categories are too long to comfortably read on their side! They are not usually used to show trends over time

Barplot

What Does It Measure?

Like column charts, bar charts are frequently used to compare the total number of items within a category, for example total sales or the number of respondents that selected a particular answer. However, they're also handy for visualizing sub-categories using colour coding.

Barplot

What Sources of Data Does It Use?

Data used to compile bar charts could come from Google Analytics, your CRM, sales figures or any other kind of database that stores data numerically.

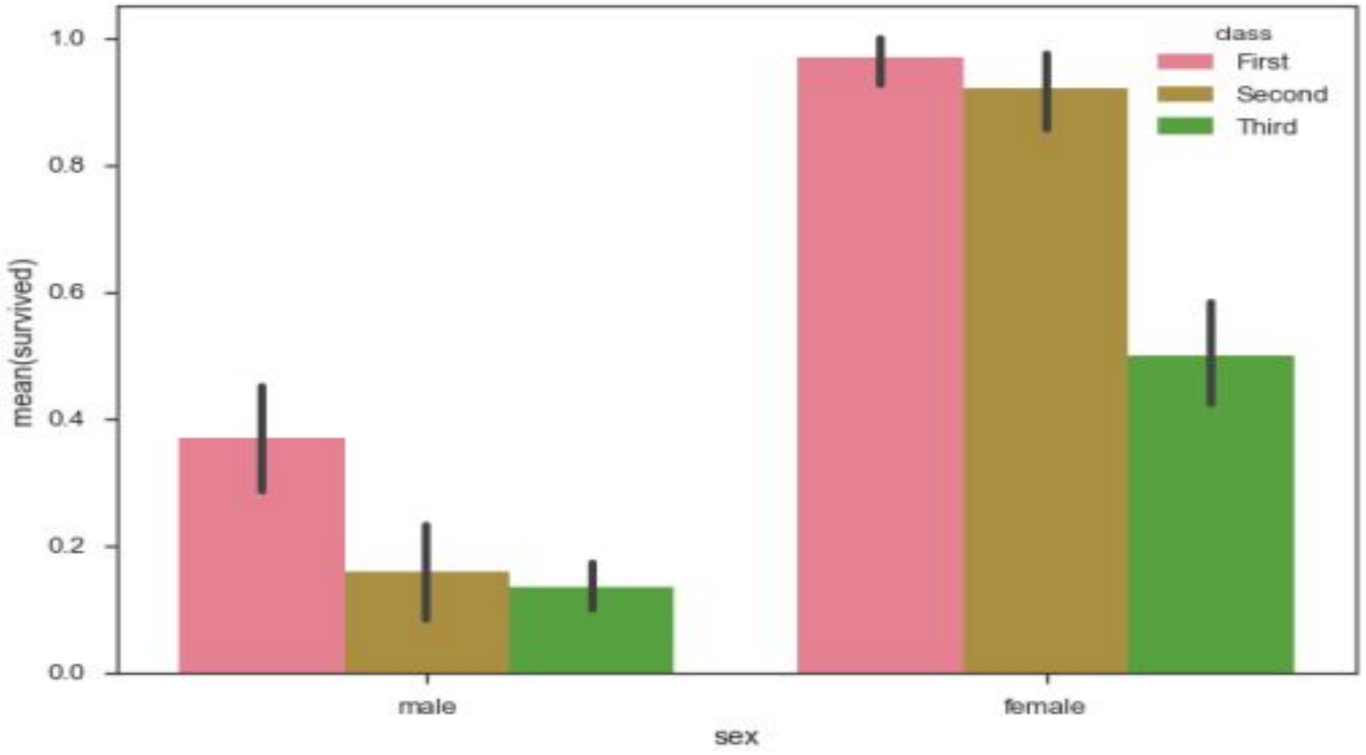
Barplot

Example

```
import pandas as pd
import seaborn as sb
from matplotlib import pyplot as plt
df = sb.load_dataset('titanic')
sb.barplot(x="sex", y="survived", hue="class", data=df)
plt.show()
```

Barplot

Output :



Boxplot

What is It?

1. The box plot, also called the box and whisker diagram is used for depicting groups of numerical data through the quartiles. It is known as the box and whisker diagram because it is composed of a box and whiskers. Boxplot is also used for detecting the outlier in a data set.

Boxplot

2. A box plot is composed of a summary of 5 different data points: the minimum, first quartile, median, third quartile, and maximum.- Minimum

- First Quartile or 25%

- Median (Second Quartile) or 50%

- Third Quartile or 75%

- Maximum

Boxplot

What Does It Visualize?

A Box and Whisker Plot (or Box Plot) is a convenient way of visually displaying **the data distribution through their quartiles**. The lines extending parallel from the boxes are known as the “whiskers”, which are used to indicate variability outside the upper and lower quartiles.

Boxplot

What Does It Measure?

A boxplot is a standardized way of displaying the distribution of data based on a five number summary (“minimum”, first quartile (Q1), median, third quartile (Q3), and “maximum”). It can tell you about your outliers and what their values are.

Boxplot

What type of data does a box plot use?

In descriptive statistics, a box plot or boxplot (also known as box and whisker plot) is a type of chart often used in exploratory data analysis. Box plots visually show the distribution of **numerical data** and skewness through displaying the data quartiles (or percentiles) and averages.

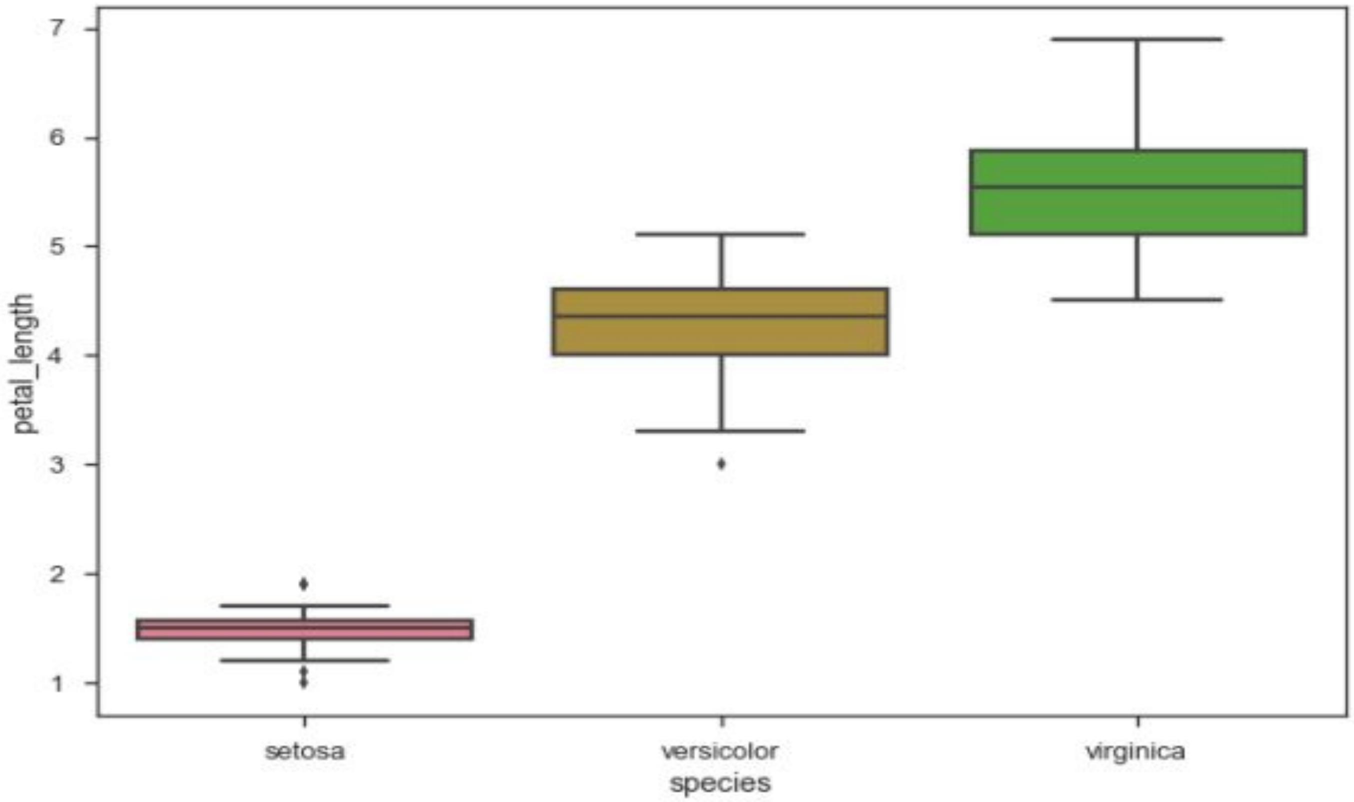
Boxplot

Example

```
import pandas as pd

import seaborn as sb

from matplotlib import pyplot as plt

df = sb.load_dataset('iris')

sb.swarm plot(x="species", y="petal_length", data=df)

plt.show()
```

Boxplot

Output :



Violinplot

What is It?

A violin plot plays a similar role as a box and whisker plot. It **shows the distribution of quantitative data across several levels of one (or more) categorical variables such that those distributions can be compared**.

Violinplot

What Does It Visualize?

A violin plot is a hybrid of a box plot and a kernel density plot, which shows peaks in the data. It is used to visualize **the distribution of numerical data**. Unlike a box plot that can only show summary statistics, violin plots depict summary statistics and the density of each variable.

Violinplot

What Does It Measure?

A violin plot **depicts distributions of numeric data for one or more groups using density curves**. The width of each curve corresponds with the approximate frequency of data points in each region. Densities are frequently accompanied by an overlaid chart type, such as box plot, to provide additional information.

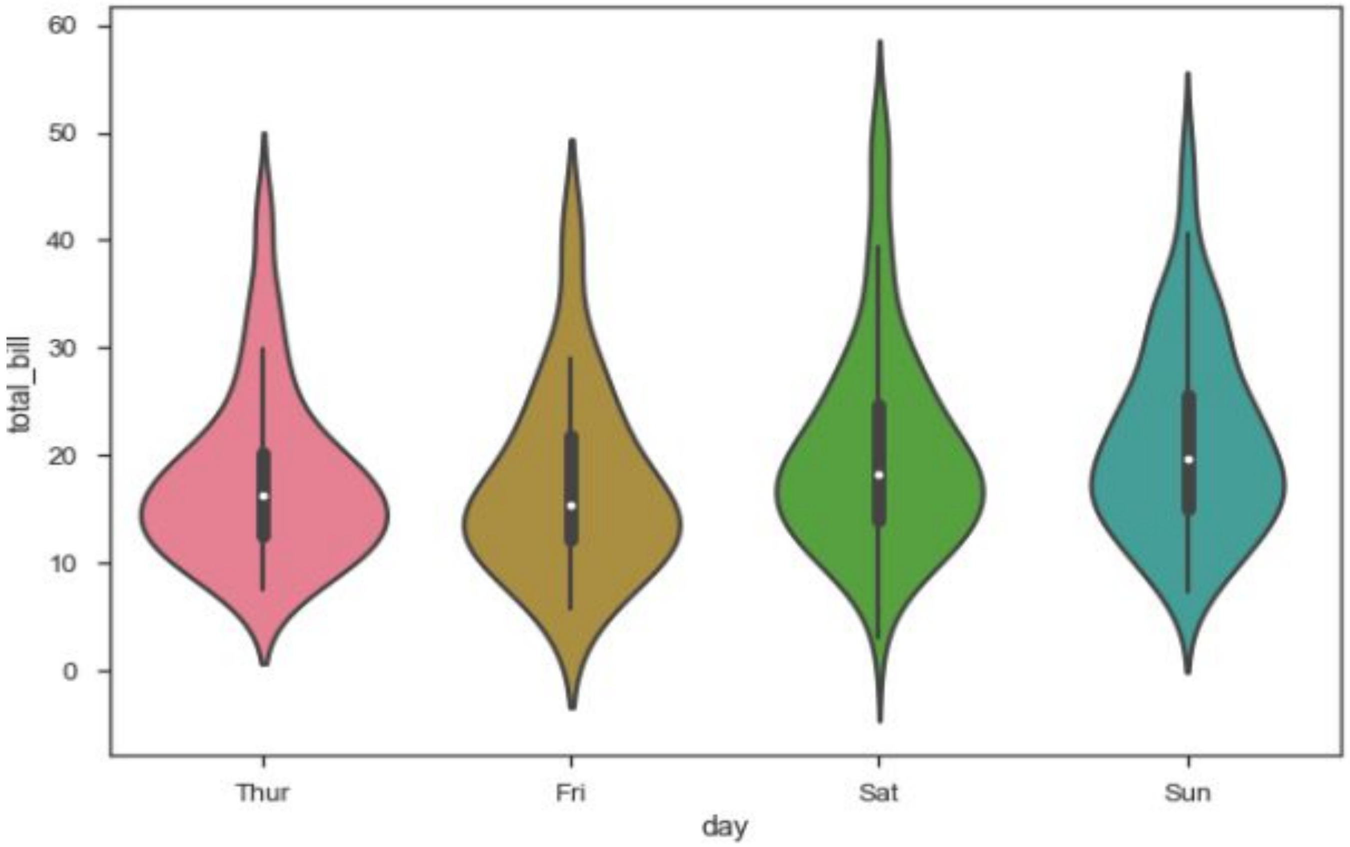
Violinplot

Example

```
 import pandas as pd
 import seaborn as sb
 from matplotlib import pyplot as plt
df = sb.load_dataset('tips')
sb.violinplot(x="day", y="total_bill", data=df)
plt.show()
```

Violinplot

Output :



Heatmap

What is It?

A heatmap **contains values representing various shades of the same colour for each value to be plotted**. Usually the darker shades of the chart represent higher values than the lighter shade. For a very different value a completely different colour can also be used.

Heatmap

What Does It Visualize?

Heatmap visualization or heatmap data visualization is **a method of graphically representing numerical data where the value of each data point is indicated using colors**.

Heatmap

What Does It Measure?

A heat map (or heatmap) is a data visualization technique that shows **magnitude of a phenomenon as color in two dimensions**. The variation in color may be by hue or intensity, giving obvious visual cues to the reader about how the phenomenon is clustered or varies over space.

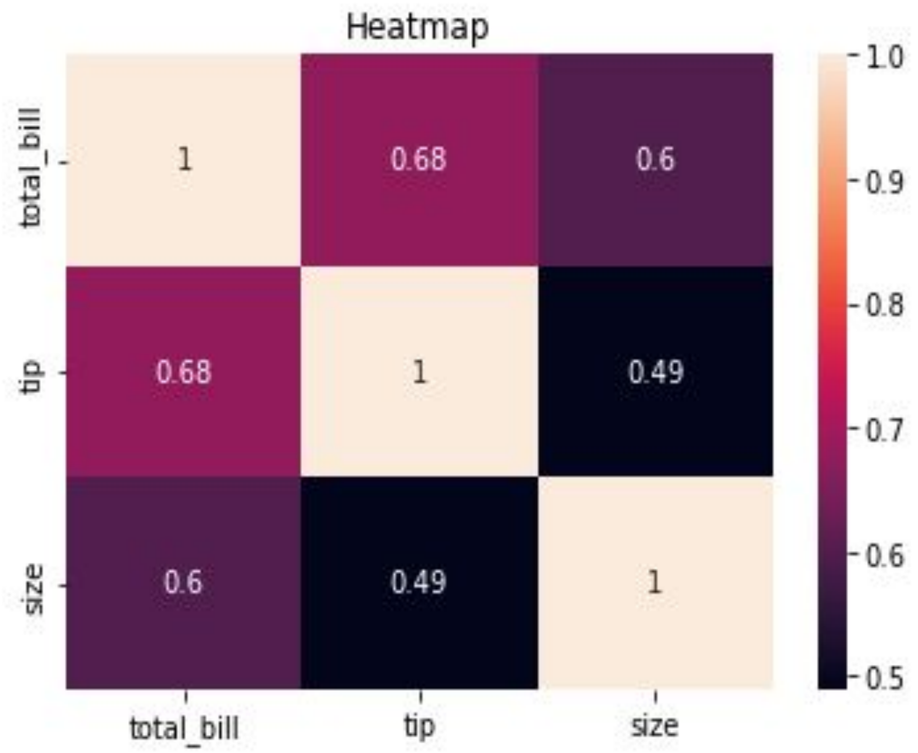
Heatmap

Example

```
 import pandas as pd

 import seaborn as sb

 from matplotlib import pyplot as plt

df = sb.load_dataset('tips')

sb.heatmap(df.corr(),annot=True)

plt.title('Heatmap')

plt.show()
```

Heatmap

Output:



Stripplot

What is It?

A strip plot is very simple to understand. It is basically **a scatter plot that differentiates different categories**. So, all the data that corresponds to each category is shown as a scatter plot, and all the observations and collected data that are visualized are shown, side-by-side on a single graph.

Stripplot

What Does It Visualize?

A strip plot is a single-axis scatter plot that is used to visualise **the distribution of many individual one-dimensional values**. The values are plotted as dots along one unique axis, and the dots with the same value can overlap.

Stripplot

What Does It Measure?

The strip-plot design is specifically suited for a two-factor experiment in which the desired precision for measuring the interaction effects between the two factors is higher than that for measuring the main effect

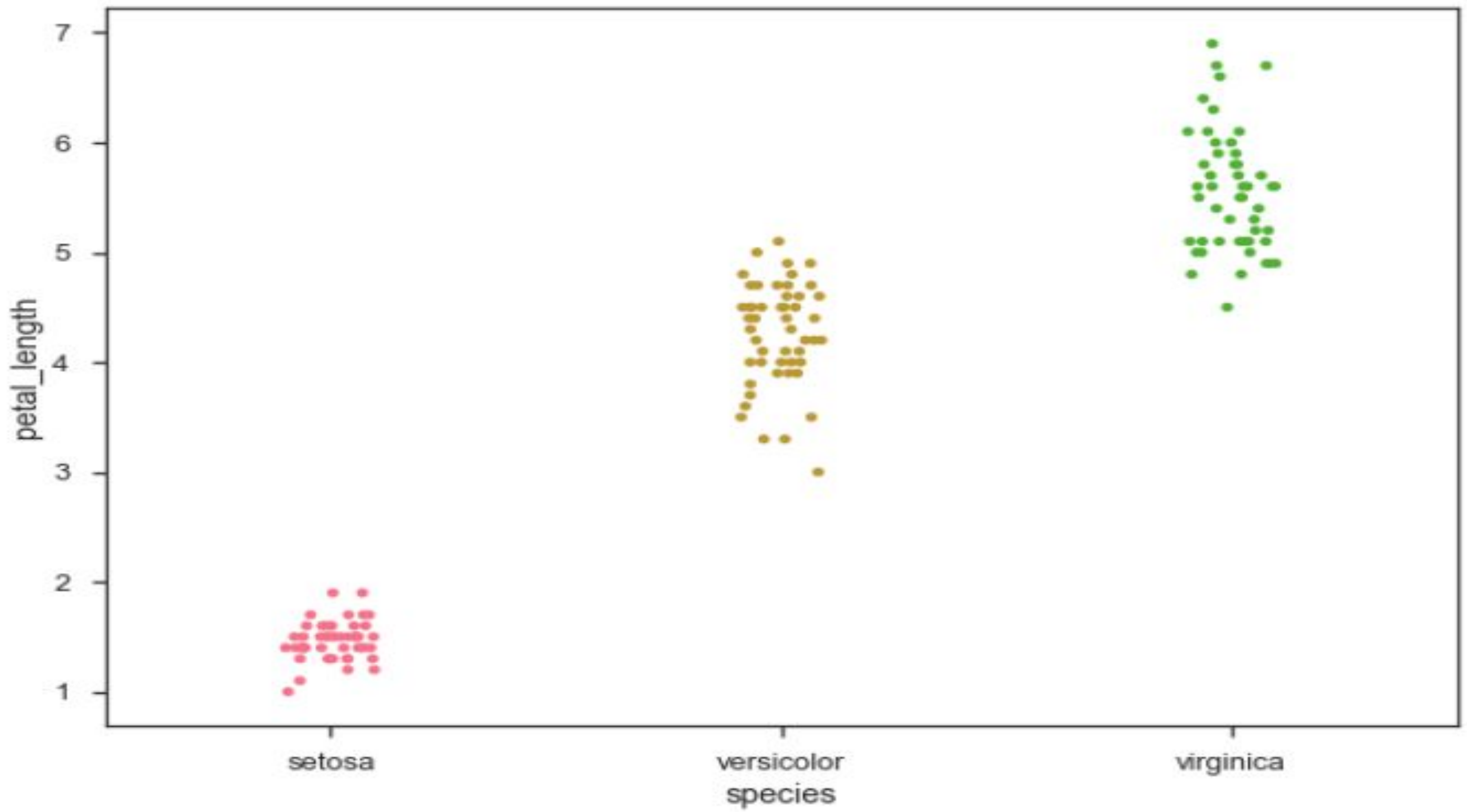
Stripplot

Example

```
import pandas as pd

import seaborn as sb

from matplotlib import pyplot as plt

df = sb.load_dataset('iris')

sb.stripplot(x="species", y="petal_length", data=df, jitter=Ture)

plt.show()
```

Stripplot

Output:



Swarmplot

What is It?

This function is **similar to stripplot() , but the points are adjusted (only along the categorical axis) so that they don't overlap**. This gives a better representation of the distribution of values, but it does not scale well to large numbers of observations. This style of plot is sometimes called a “beeswarm”.

Sarmplot

What Does It Visualize?

A swarmplot **shows all the data points** and that helps to understand the distribution in a better manner. It also helps to understand how the data is distributed across a categorical attribute and how the continuous variable is varying within a category.

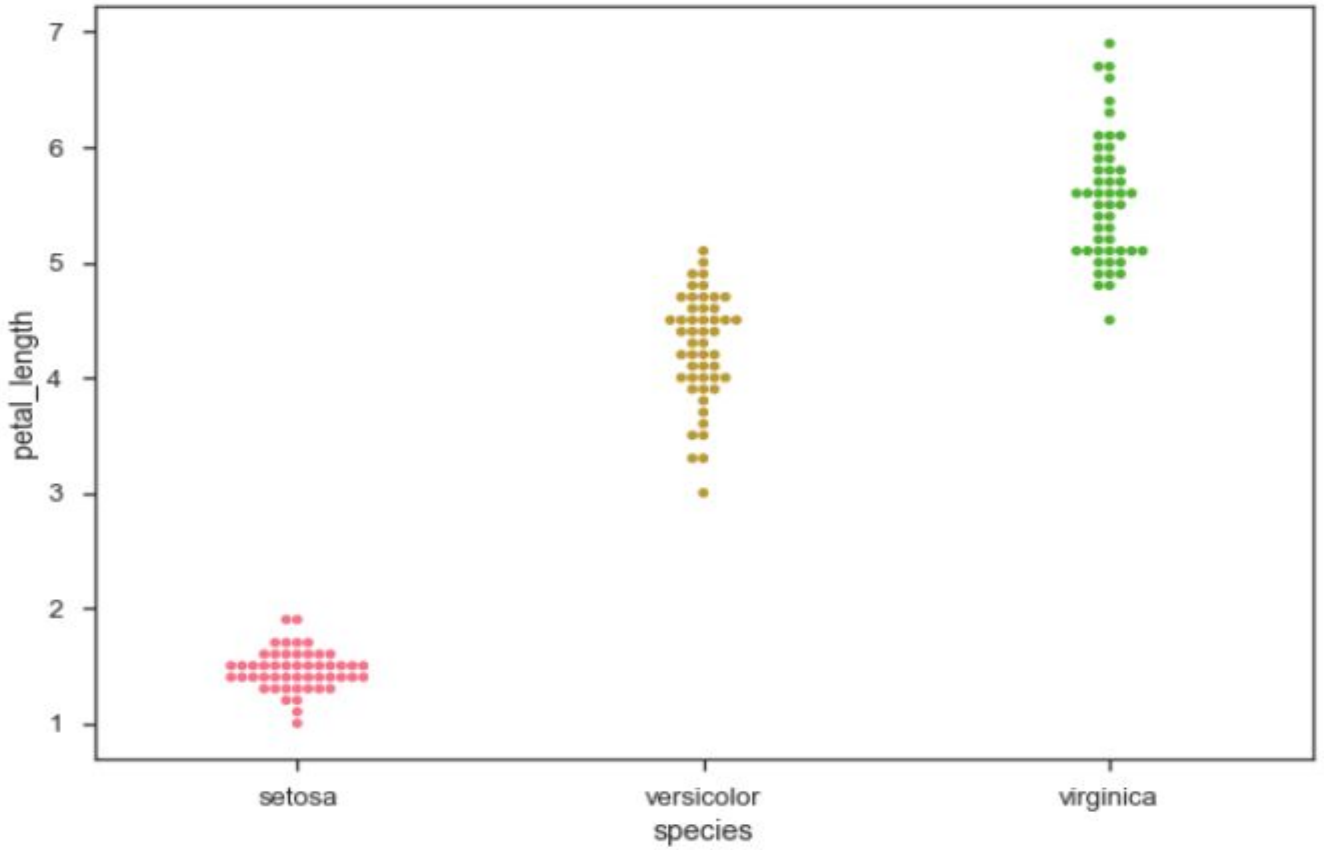
Swarmplot

Example

```
import pandas as pd

import seaborn as sb

from matplotlib import pyplot as plt

df = sb.load_dataset('iris')

sb.swarmplot(x="species", y="petal_length", data=df)

plt.show()
```

Swarmplot

Output:



Lineplot

What is it?

Draw a line plot with possibility of several semantic groupings. **The relationship between x and y can be shown for different subsets of the data using the hue , size , and style parameters**. These parameters control what visual semantics are used to identify the different subsets.

Lineplot

What Does It Visualize?

These are super simple and very popular, because they give you an immediate idea of how a trend emerged over time. You can see when peaks and troughs hit, whether the overall values are going up or down, and when there’s a sharp spike or drop in numbers.

Lineplot

What Does It Measure?

There are many different business cases that work well with line charts. Pretty much anything that compares data, or shows changes, over time is well suited to this type of visualization. Again, it’s all about visualizing a trend. You can also compare changes over the same period of time for more than one group or category very easily, by adding a “break by” category

Lineplot

What Sources of Data Does It Use?

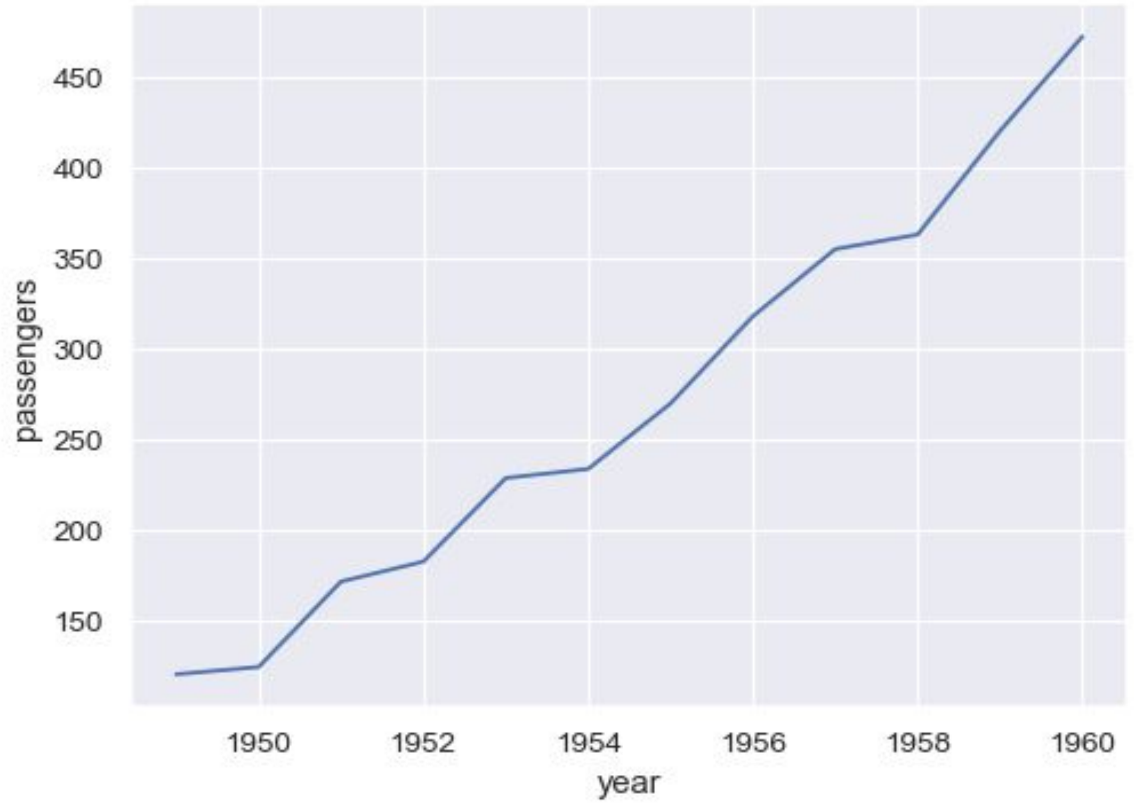
Again, anything that gives solid, discrete numbers, organized by time.

Lineplot

```
import seaborn as sns

flights = sns.load_dataset("flights")

may_flights = flights.query("month == 'May'")

sns.lineplot(data=may_flights, x="year", y="passengers")

plt.show()
```

Lineplot

Output:



lmplot

What is It?

lmplot() method is used **to draw a scatter plot onto a FacetGrid**. Parameters : This method is accepting the following parameters that are described below: x, y: (optional) This parameters are column names in data.

What Does It Visualize?

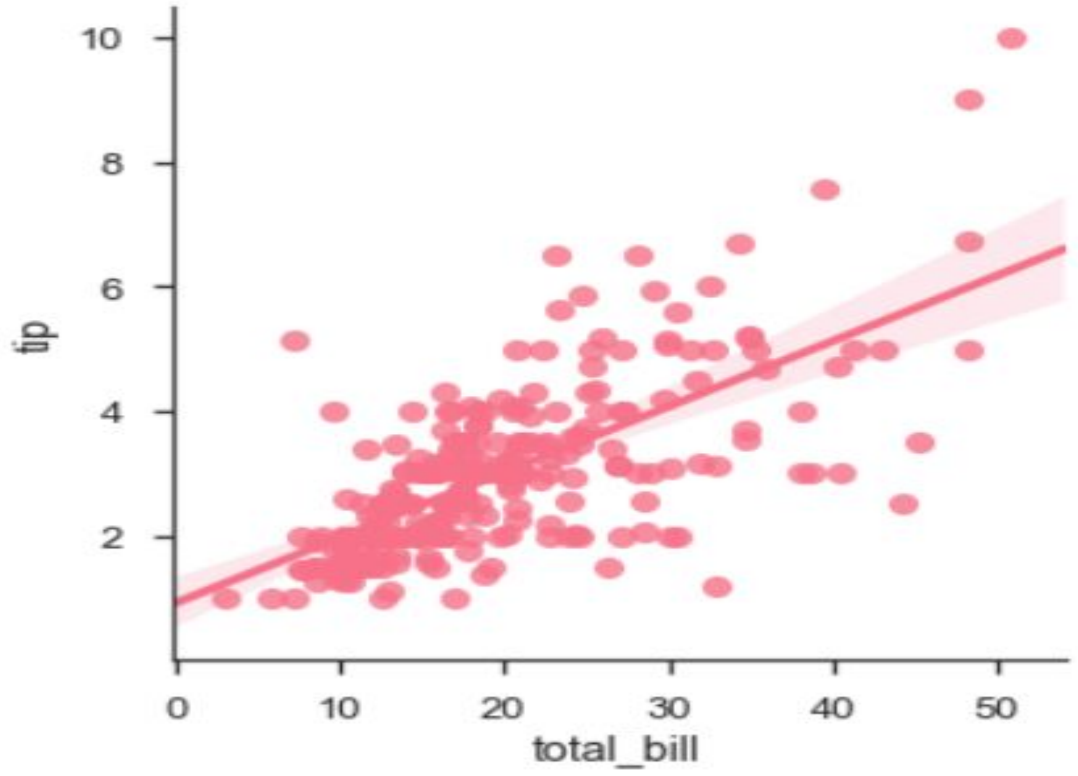
Visualize linear relationship between two different variable

lmplot

```
import pandas as pd

import seaborn as sb

from matplotlib import pyplot as plt

df = sb.load_dataset('tips')

sb.regplot(x="total_bill", y="tip", data=df)

sb.lmplot(x="total_bill", y="tip", data=df)

plt.show( )
```

lmplot

Output :



Pairplot

What is it?

Pairplot **visualizes given data to find the relationship between them where the variables can be continuous or categorical**. Plot pairwise relationships in a data-set. Pairplot is a module of seaborn library which provides a high-level interface for drawing attractive and informative statistical graphics.

Pairplot

What Does It Visualize?

Pairplot **visualizes given data to find the relationship between them where the variables can be continuous or categorical**. Plot pairwise relationships in a data-set. Pairplot is a module of seaborn library which provides a high-level interface for drawing attractive and informative statistical graphics.

Pairplot

What Does It Measure?

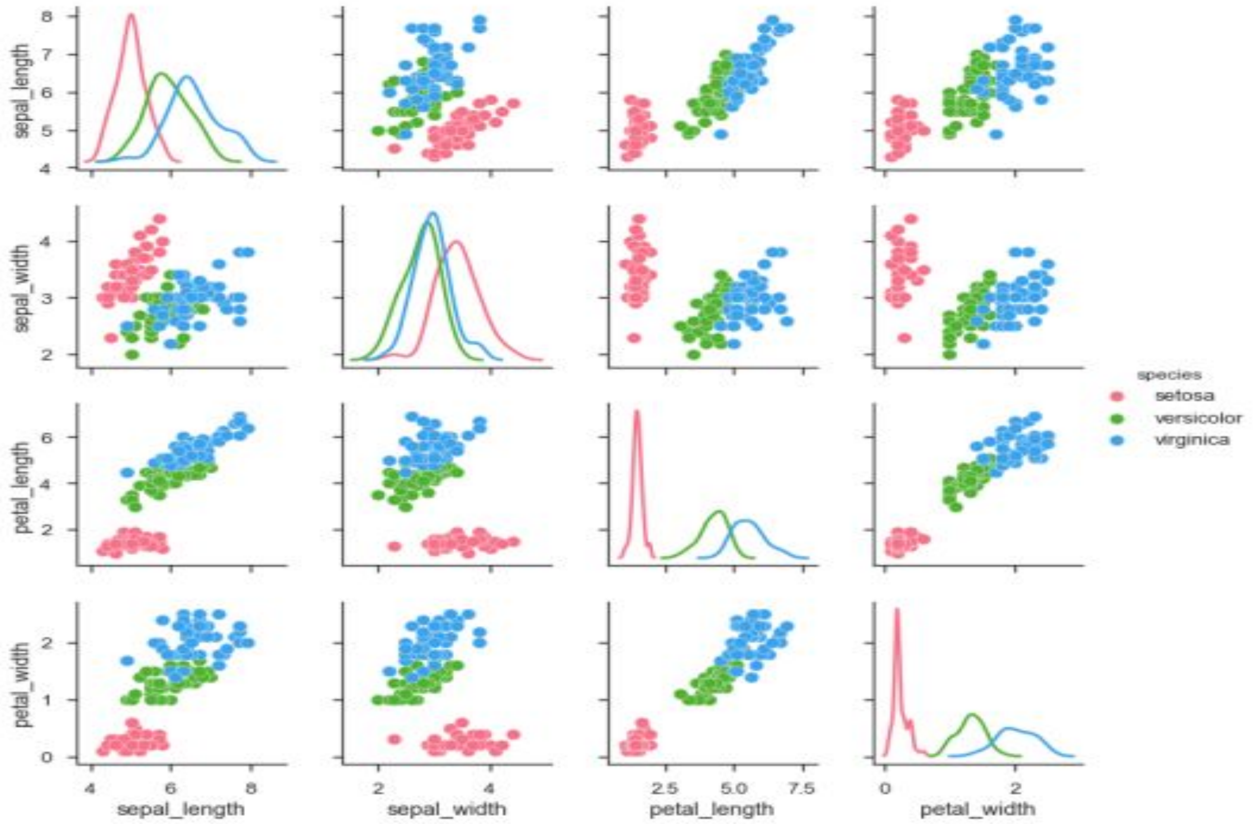
A pairplot **plot a pairwise relationships in a dataset**. The pairplot function creates a grid of Axes such that each variable in data will by shared in the y-axis across a single row and in the x-axis across a single column.

Pairplot

```
import seaborn as sb

from matplotlib import pyplot as plt

df = sb.load_dataset('iris')

sb.set_style("ticks")
sb.pairplot(df,hue='species',diag_kind="kde",kind="scatter",palette="husl")
plt.show()
```

Pairplot

Output :



The End