Scatter Plot

**Scatter plots** are the graphs that present the relationship between two variables in a data-set. It represents data points on a two-dimensional plane or on a **Cartesian system**. The independent variable or attribute is plotted on the X-axis, while the dependent variable is plotted on the Y-axis. These plots are often called **scatter graphs** or **scatter diagrams**.

|  |
| --- |
| * [Bar Graph](https://byjus.com/maths/bar-graph/) * [Graphical Representation](https://byjus.com/maths/graphical-representation/) * [Correlation](https://byjus.com/maths/correlation/) * [Data Sets](https://byjus.com/maths/data-sets/) |

Scatter plot Graph

A scatter plot is also called a scatter chart, scattergram, or scatter plot, XY graph. The scatter diagram graphs numerical data pairs, with one variable on each axis, show their relationship. Now the question comes for everyone: **when to use a scatter plot**?

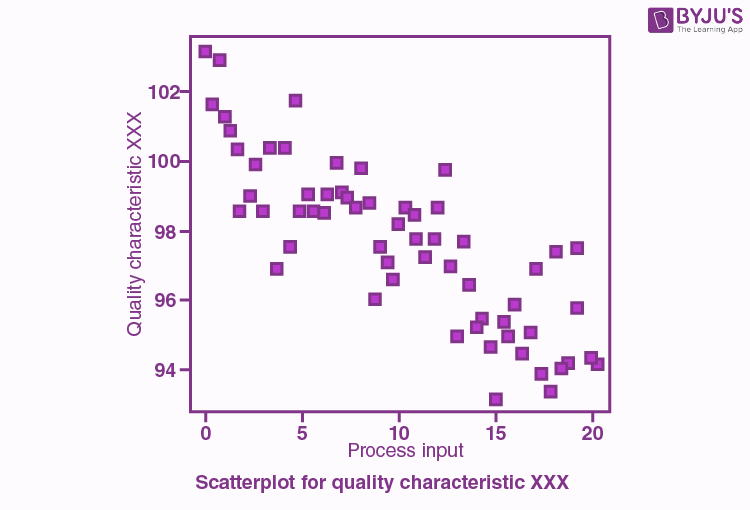
Scatter plots are used in either of the following situations.

* When we have paired numerical data
* When there are multiple values of the dependent variable for a unique value of an independent variable
* In determining the relationship between variables in some scenarios, such as identifying potential root causes of problems, checking whether two products that appear to be related both occur with the exact cause and so on.

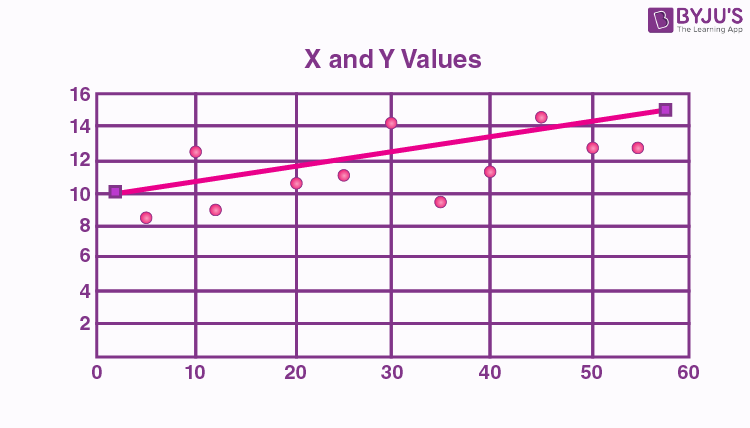
Scatter Plot Uses and Examples

Scatter plots instantly report a large volume of data. It is beneficial in the following situations –

* For a large set of data points given
* Each set comprises a pair of values
* The given data is in numeric form



The line drawn in a scatter plot, which is near to almost all the points in the plot is known as “**line of best fit**” or “**trend line**“. See the graph below for an example.



Scatter plot Correlation

We know that the correlation is a statistical measure of the relationship between the two variables’ relative movements. If the variables are correlated, the points will fall along a line or curve. The better the correlation, the closer the points will touch the line. This cause examination tool is considered as one of the seven essential quality tools.

Types of correlation

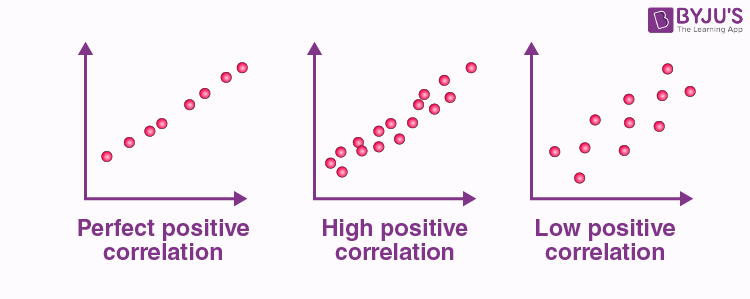
The scatter plot explains the correlation between two attributes or variables. It represents how closely the two variables are connected. There can be three such situations to see the relation between the two variables –

1. Positive Correlation
2. Negative Correlation
3. No Correlation

Positive Correlation

When the points in the graph are rising, moving from left to right, then the scatter plot shows a positive correlation. It means the values of one variable are increasing with respect to another. Now positive correlation can further be classified into three categories:

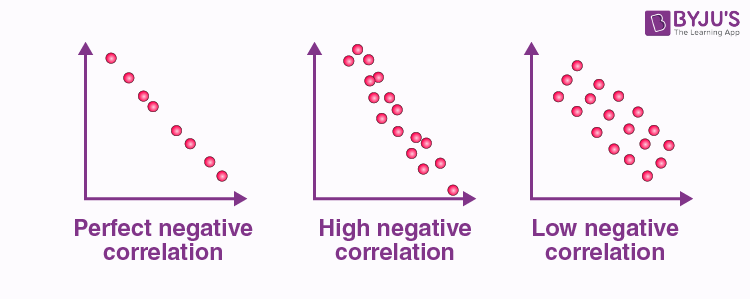
* **Perfect Positive**– Which represents a perfectly straight line
* **High Positive** – All points are nearby
* **Low Positive** – When all the points are scattered



Negative Correlation

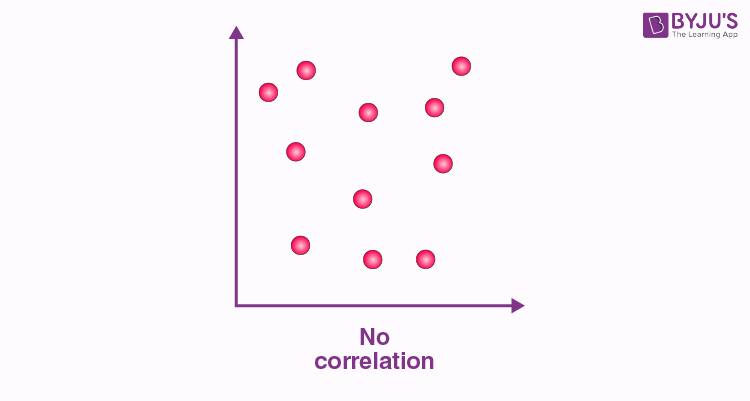
When the points in the scatter graph fall while moving left to right, then it is called a negative correlation. It means the values of one variable are decreasing with respect to another. These are also of three types:

* **Perfect Negative** – Which form almost a straight line
* **High Negative** – When points are near to one another
* **Low Negative** – When points are in scattered form



No Correlation

When the points are scattered all over the graph and it is difficult to conclude whether the values are increasing or decreasing, then there is no correlation between the variables.



Scatter plot Example

Let us understand how to construct a scatter plot with the help of the below example.

**Question:**

Draw a scatter plot for the given data that shows the number of games played and scores obtained in each instance.

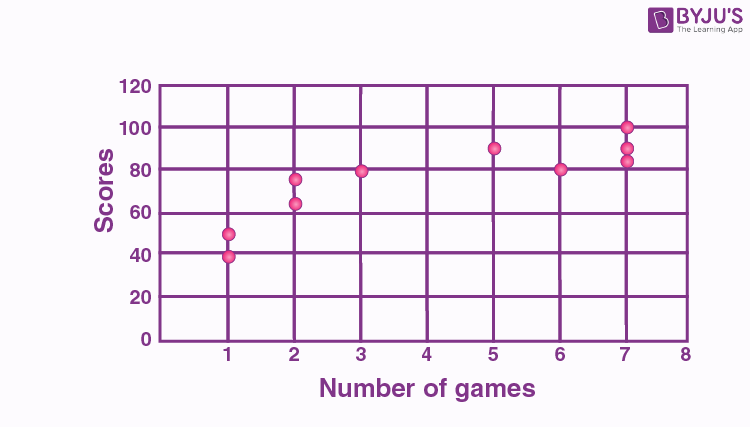
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| No. of games | 3 | 5 | 2 | 6 | 7 | 1 | 2 | 7 | 1 | 7 |
| Scores | 80 | 90 | 75 | 80 | 90 | 50 | 65 | 85 | 40 | 100 |

**Solution:**

X-axis or horizontal axis: Number of games

Y-axis or vertical axis: Scores

Now, the scatter graph will be:



Drawing A Scatter Plot Using Pandas DataFrame

1. [Home](https://pythontic.com/)
2. [Pandas](https://pythontic.com/pandas)
3. [Dataframe-plotting](https://pythontic.com/pandas/dataframe-plotting)
4. Scatter Plot

Overview:

* A **scatter plot**is a diagram drawn between two distributions of variables X and Y on a two dimensional plane.
* **Scatter plot** is used as an initial screening tool while analyzing two variables for any relationship (linear, non-linear, inverse relationships) that may exist between them.
* A scatter plot is used only as an **initial tool**in the process of finding any relationship between two variables. Even if a relationship is found between two variables using scatter plot, it may not be true that one variable influences another variable. To establish relationship between two variables tools like [correlation](https://pythontic.com/visualization/charts/correlation) can be used.

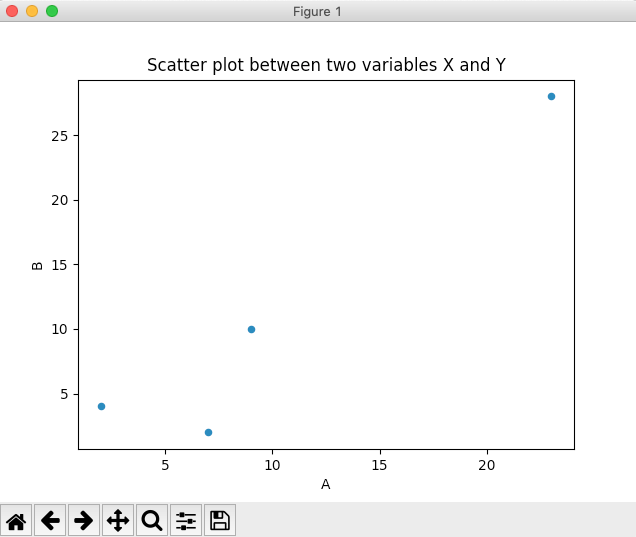
Plotting a scatter plot using Pandas DataFrame:

* The pandas DataFrame class in Python has a member plot.
* Invoking the **scatter()** method on the plot member draws a **scatter plot** between two given columns of a pandas DataFrame.
* A pandas DataFrame can have several columns. Any two columns can be chosen as X and Y parameters for the **scatter()** method.

Example 1:

|  |
| --- |
| # Example Python program to draw a scatter plot  # for two columns of a pandas DataFrame  import pandas as pd  import matplotlib.pyplot as plot    # List of tuples  data = [(2, 4),          (23, 28),          (7, 2),          (9, 10)]    # Load data into pandas DataFrame  dataFrame = pd.DataFrame(data=data, columns=['A','B']);    # Draw a scatter plot  dataFrame.plot.scatter(x='A', y='B', title= "Scatter plot between two variables X and Y");  plot.show(block=True); |

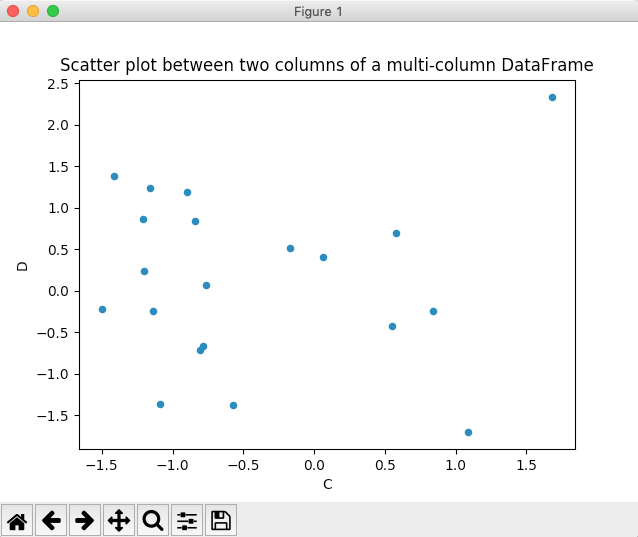
Output:



Example 2:

|  |
| --- |
| # Example Python program to draw a scatter plot  # for two columns of a multi-column DataFrame  import pandas as pd  import numpy as np  import matplotlib.pyplot as plot    # Create an ndarray with three columns and 20 rows  data = np.random.randn(20, 4);    # Load data into pandas DataFrame  dataFrame = pd.DataFrame(data=data, columns=['A', 'B', 'C', 'D']);    # Draw a scatter plot  dataFrame.plot.scatter(x='C', y='D', title= "Scatter plot between two columns of a multi-column DataFrame");  plot.show(block=True); |

Output:



[area-plot](https://pythontic.com/pandas/dataframe-plotting/area-plot)

[bar chart](https://pythontic.com/pandas/dataframe-plotting/bar%20chart)

[box and whisker plot](https://pythontic.com/pandas/dataframe-plotting/box%20and%20whisker%20plot)

[hexagonal binning plot](https://pythontic.com/pandas/dataframe-plotting/hexagonal%20binning%20plot)

[histogram-for-each-column](https://pythontic.com/pandas/dataframe-plotting/histogram-for-each-column)

[Kernel Density Estimation plot](https://pythontic.com/pandas/dataframe-plotting/kernel%20density%20estimation%20plot)

[line chart](https://pythontic.com/pandas/dataframe-plotting/line%20chart)

Copyright 2022 © pythontic.com

**Correlation**

When the two sets of data are strongly linked together we say they have a **High Correlation**.

The word Correlation is made of **Co-** (meaning "together"), and **Relation**

* Correlation is **Positive** when the values **increase** together, and
* Correlation is **Negative** when one value **decreases** as the other increases

# Like this: Pandas Scatter Plot – DataFrame.plot.scatter()

* Last Updated : 25 Feb, 2021

A Scatter plot is a type of data visualization technique that shows the relationship between two numerical variables. For plotting to scatter plot using pandas there is DataFrame class and this class has a member called plot. Calling the scatter() method on the plot member draws a plot between two variables or two columns of pandas DataFrame.

***Syntax:****DataFrame.plot.scatter(x, y, s = none, c = none)*

***Parameter:***

***x:****column name to be used as horizontal coordinates for each point****y:****column name to be used as vertical coordinates for each point****s:****size of dots****c:****color of dots*

**Steps:**

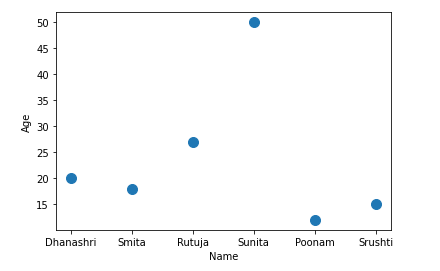
* Import necessary libraries.
* Prepare a data
* Convert prepared data into DataFrame
* Draw a scatter plot

**Example 1:**In this example, we will plot the scatter plot using dataframe, Here we will create the dataframe and plot the scatter plot using different columns.

* Python3

|  |
| --- |
| # Program to draw scatter plot using Dataframe.plot  # Import libraries  import pandas as pd    # Prepare data  data={'Name':['Dhanashri', 'Smita', 'Rutuja',                'Sunita', 'Poonam', 'Srushti'],        'Age':[20, 18, 27, 50, 12, 15]}    # Load data into DataFrame  df = pd.DataFrame(data = data);    # Draw a scatter plot  df.plot.scatter(x = 'Name', y = 'Age', s = 100); |

**Output:**



(Learn [More About Correlation](https://www.mathsisfun.com/data/correlation.html))

**Negative Correlation**

Correlations can be negative, which means there **is** a correlation but one value goes down as the other value increases.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Example : Birth Rate vs Income  The birth rate tends to be **lower** in richer countries.    Below is a scatter plot for about 100 different countries. | |  |  |  | | --- | --- | --- | | Country | Yearly Production per Person | Birth Rate | | Madagascar | $800 | 5.70 | | India | $3,100 | 2.85 | | Mexico | $9,600 | 2.49 | | Taiwan | $25,300 | 1.57 | | Norway | $40,000 | 1.78 | |