

How can I create a scatter matrix in Pandas? Can you provide some examples?

Authored by
stats writer

May 5, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I create a scatter matrix in Pandas? Can you provide some examples?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=142851>

Creating a scatter matrix in Pandas involves using the "scatter_matrix" function from the Pandas library. This function allows for the visualization of the relationship between multiple numerical variables in a dataset through the use of scatter plots. To create a scatter matrix, the user needs to pass a pandas DataFrame as the parameter, along with other optional parameters such as the figure size and labels. The resulting scatter matrix will display the scatter plots for each combination of variables in the DataFrame, providing a comprehensive view of the data. Examples of creating scatter matrices can be found in the Pandas documentation or through online resources.

Create a Scatter Matrix in Pandas (With Examples)

A scatter matrix is exactly what it sounds like - a matrix of scatterplots.

This type of matrix is useful because it allows you to visualize the relationship between multiple variables in a dataset at once.

You can use the scatter_matrix() function to create a scatter matrix from a pandas DataFrame:

```
pd.plotting.scatter_matrix(df)
```

The following examples show how to use this syntax in practice with the following pandas DataFrame:

```
import pandas as pd  
import numpy as np
```

```
#make this example reproducible
```

```
np.random.seed(0)
```

```
#create DataFrame
```

```
df = pd.DataFrame({'points': np.random.randn(1000),
```

```
'assists': np.random.randn(1000),
```

```
'rebounds': np.random.randn(1000)})
```

```
#view first five rows of DataFrame
```

```
df.head()
```

```
points assists rebounds
```

```
0 1.764052 0.555963 -1.532921
```

```
1 0.400157 0.892474 -1.711970
```

```
2 0.978738 -0.422315 0.046135
```

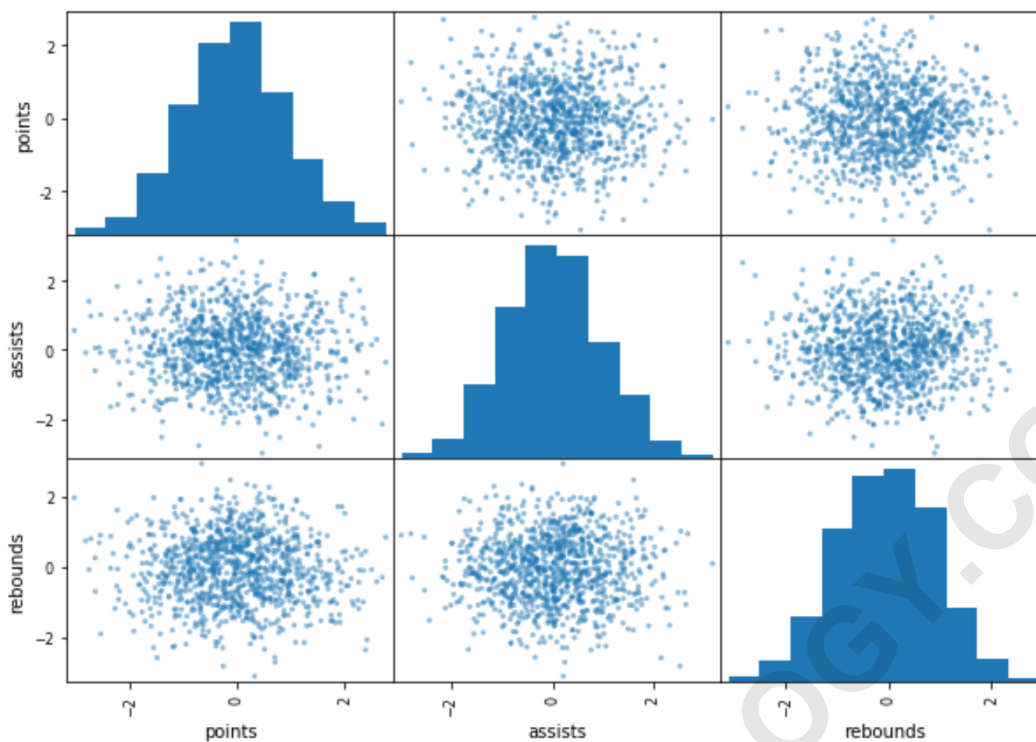
```
3 2.240893 0.104714 -0.958374
```

```
4 1.867558 0.228053 -0.080812
```

Example 1: Basic Scatter Matrix

The following code shows how to create a basic scatter matrix:

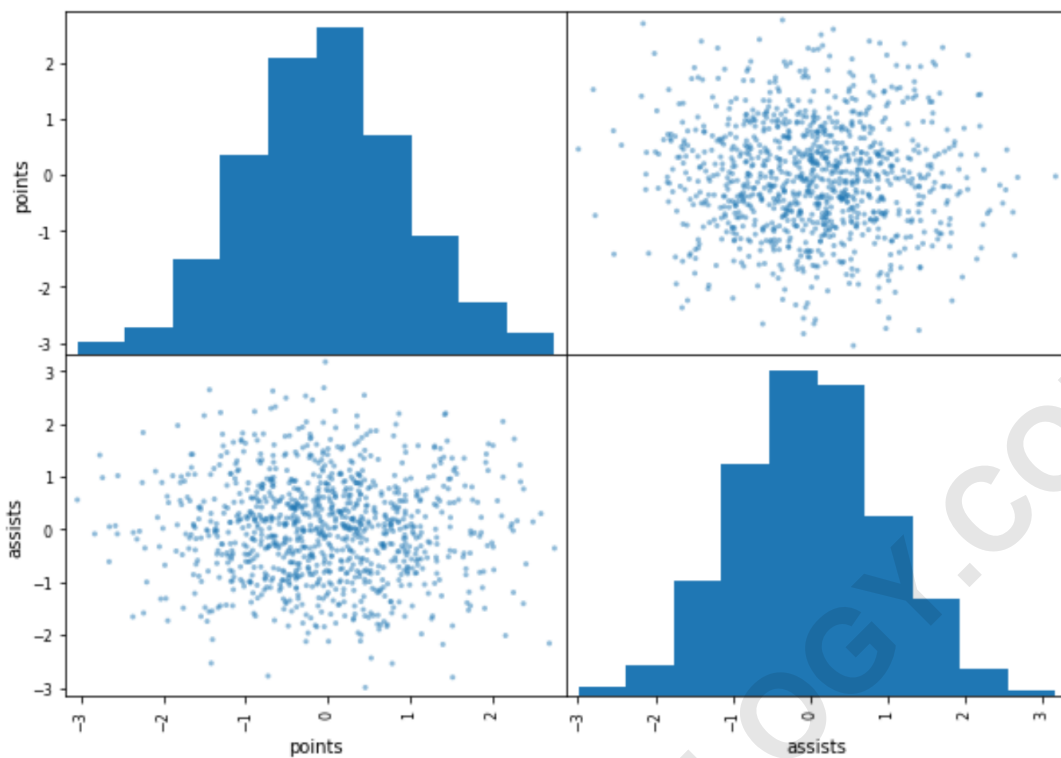
```
pd.plotting.scatter_matrix(df)
```



Example 2: Scatter Matrix for Specific Columns

The following code shows how to create a scatter matrix for just the first two columns in the DataFrame:

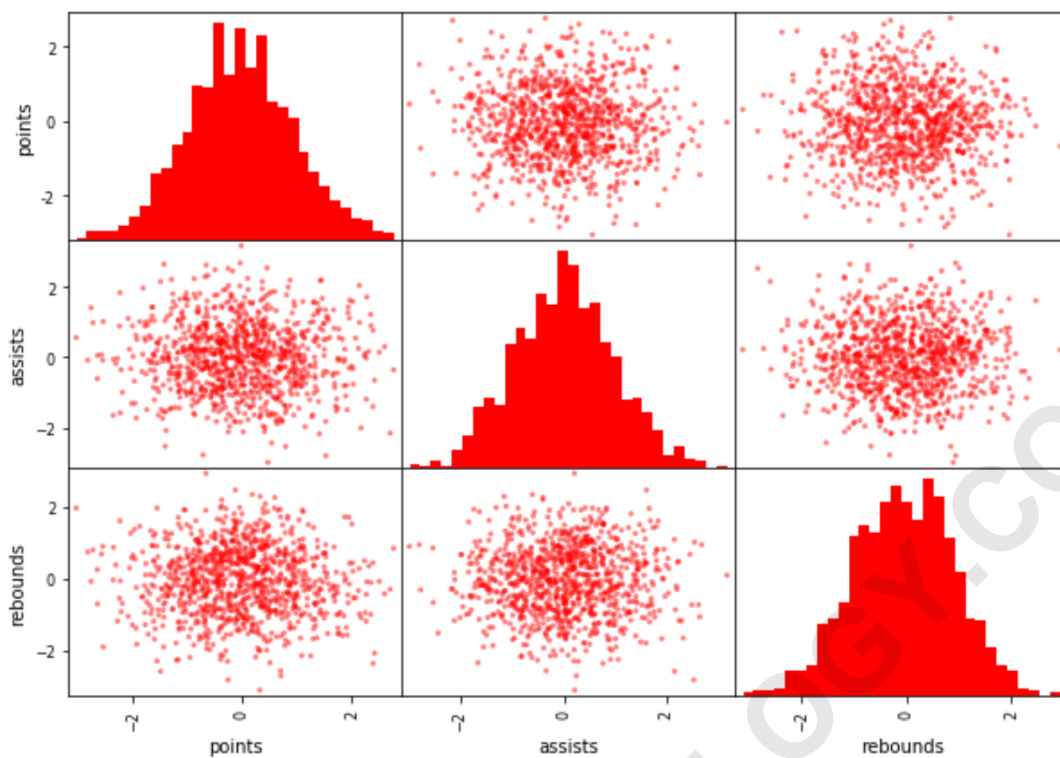
```
pd.plotting.scatter_matrix(df.iloc)
```



Example 3: Scatter Matrix with Custom Colors & Bins

The following code shows how to create a scatter matrix with custom colors and a specific number of bins for the histograms:

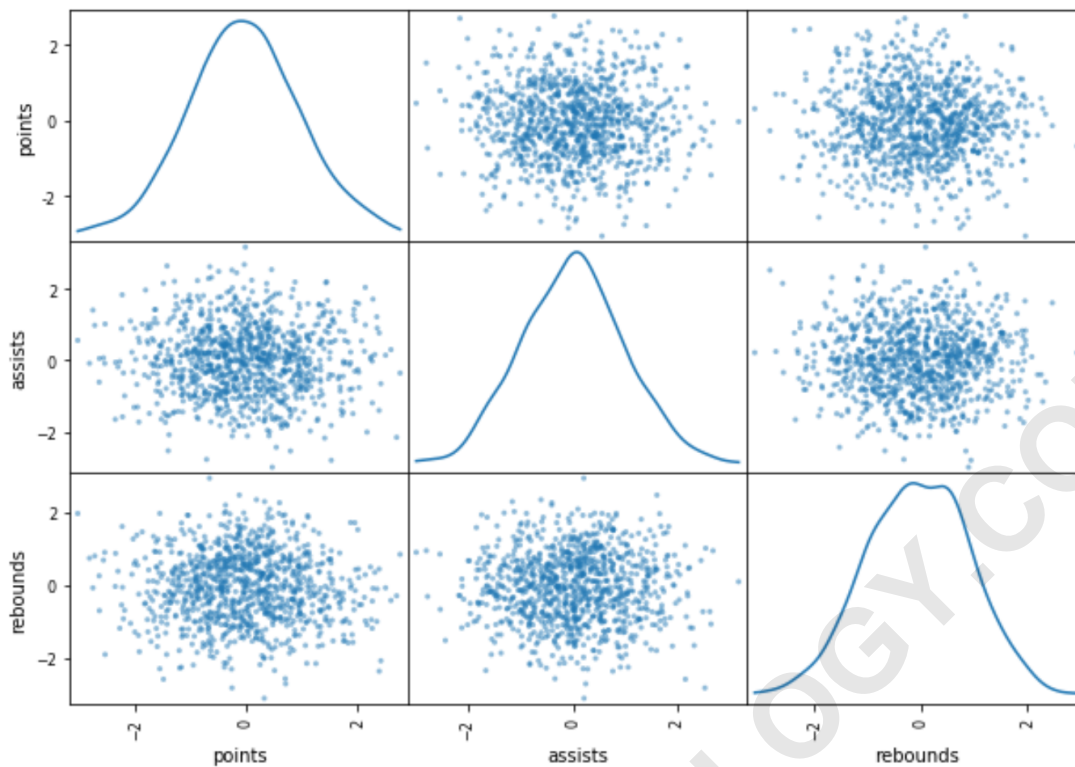
```
pd.plotting.scatter_matrix(df, color='red',  
hist_kwds={'bins':30, 'color':'red'})
```



Example 4: Scatter Matrix with KDE Plot

The following code shows how to create a scatter matrix with a kernel density estimate plot along the diagonals of the matrix instead of a histogram:

```
pd.plotting.scatter_matrix(df, diagonal='kde')
```



You can find the complete online documentation for the `scatter_matrix()` function .

The following tutorials explain how to create other common charts in Python: