How can I create a Seaborn scatterplot with a correlation coefficient?

Authored by

stats writer

June 24, 2024

A Seaborn scatterplot can be created with a correlation coefficient by using the built-in "regplot" function. This function allows for plotting a scatterplot with a linear regression line and displaying the correlation coefficient on the graph. Additionally, the "scatterplot" function can be used to plot a basic scatterplot and the correlation coefficient can be calculated separately and added to the graph as a text annotation. Both methods provide a visual representation of the relationship between two variables and the strength of their correlation.

Create Seaborn Scatterplot with Correlation Coefficient

You can use the following basic syntax to create a scatterplot in seaborn and add a to the plot:

```
import scipy
import matplotlib.pyplotas plt
import seaborn as sns

#calculate correlation coefficient between x and y
r = scipy.stats.pearsonr(x=df.x, y=df.y)

#create scatterplot
sns.scatterplot(data=df, x=df.x, y=df.y)

#add correlation coefficient to plot
plt.text(5, 30, 'r = ' + str(round(r, 2)))
```

The following example shows how to use this syntax in practice.

Example: Create Seaborn Scatterplot with Correlation Coefficient

Suppose we have the following pandas DataFrame that shows the points and assists for various basketball players:

```
import pandas as pd

#create DataFrame
df = pd.DataFrame({'team': ,
'points': ,
'assists': })

#view DataFrame
print(df)
```

```
team points assists
0 A 12 4
1 A 11 7
2 A 18 7
3 A 15 8
4 B 14 9
5 C 20 10
6 C 25 10
7 C 24 12
```

8 D 32 10

9 D 30 15

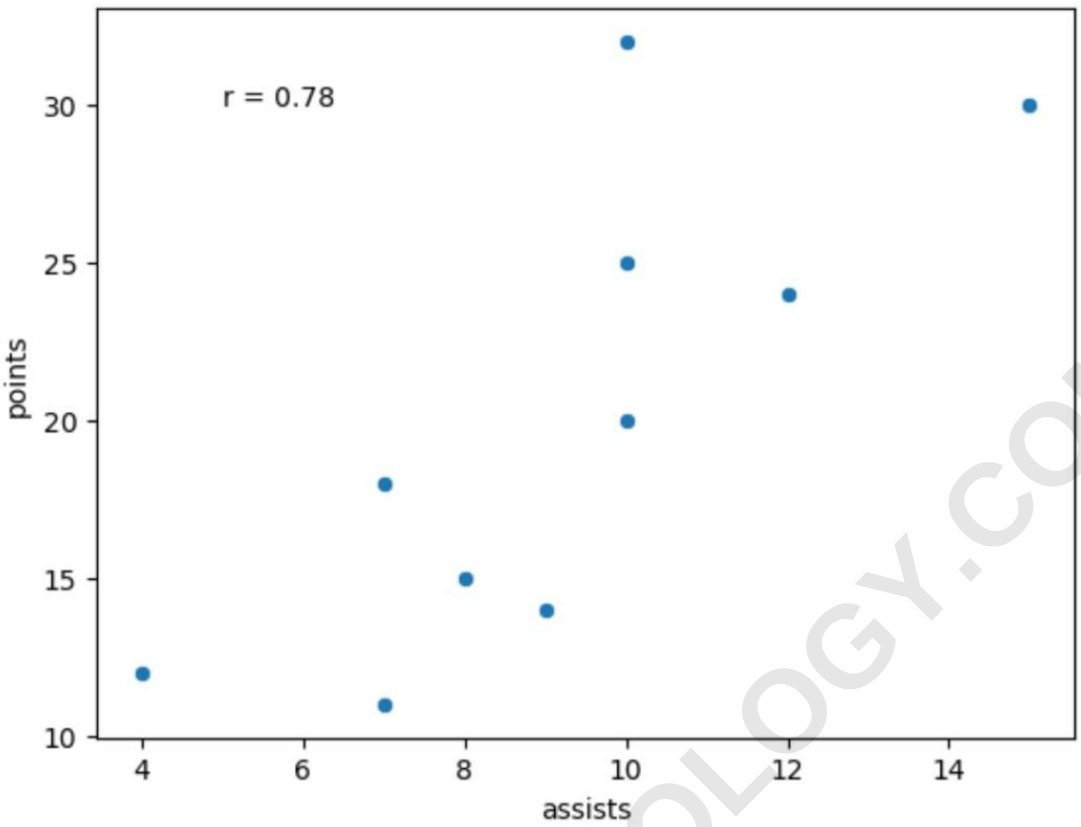
We can use the following syntax to create a scatterplot to visualize the relationship between assists and points and also use the pearsonr() function from scipy to calculate the correlation coefficient between these two variables:

```
import scipy
import matplotlib.pyplotas plt
import seaborn as sns

#calculate correlation coefficient between assists and points
r = scipy.stats.pearsonr(x=df.assists, y=df.points)

#create scatterplot
sns.scatterplot(data=df, x=df.assists, y=df.points)

#add correlation coefficient to plot
plt.text(5, 30, 'r = ' + str(round(r, 2)))
```



From the output we can see that the Pearson correlation coefficient between assists and points is 0.78.

Note that we used the round() function to round the correlation coefficient to two decimal places.

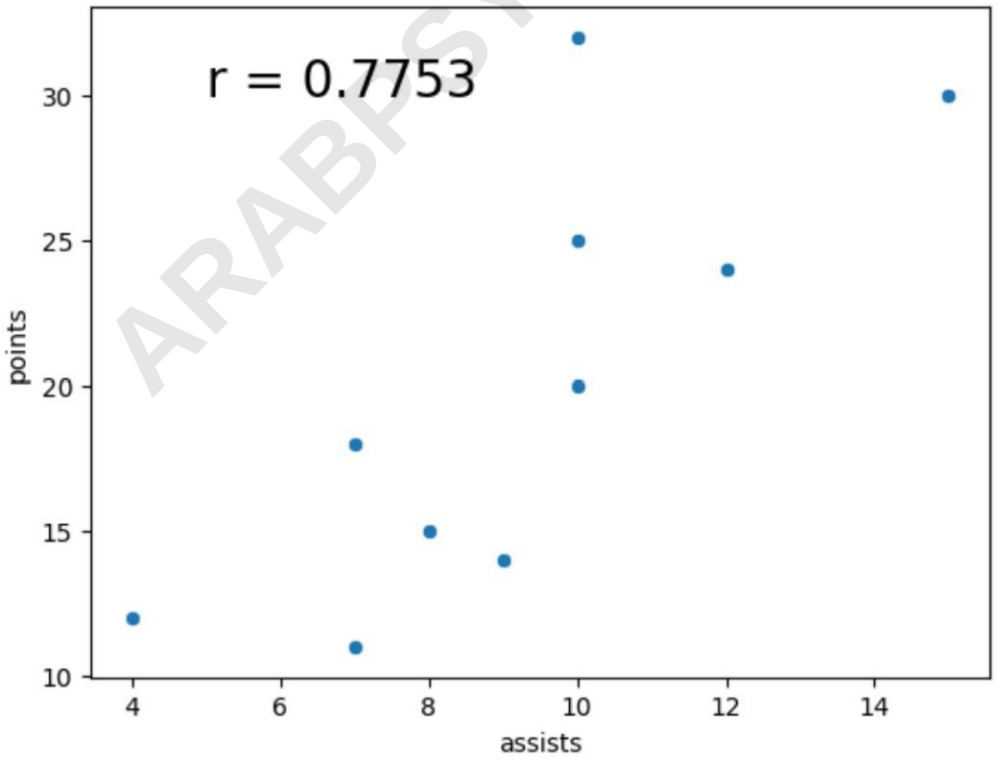
Feel free to round to a different number of decimal places and also feel free to use the fontsize argument to change the font size of the correlation coefficient on the plot:

```
import scipy
import matplotlib.pyplotas plt
import seaborn as sns

#calculate correlation coefficient between assists and points
r = scipy.stats.pearsonr(x=df.assists, y=df.points)

#create scatterplot
sns.scatterplot(data=df, x=df.assists, y=df.points)

#add correlation coefficient to plot
plt.text(5, 30, 'r = ' + str(round(r, 4)), fontsize=20))
```



Notice that the correlation coefficient is now rounded to four decimal places and the font size is much larger than the previous example.

Note: You can find the complete documentation for the seaborn scatterplot() function .

The following tutorials explain how to perform other common functions in seaborn: