The term bivariate analysis refers to the analysis of two variables. we can remember this because the prefix “bi” means “two.”

The purpose of bivariate analysis is to understand the relationship between two variables

There are three common ways to perform bivariate analysis:

1. Scatterplots

2. Correlation Coefficients

3. Simple Linear Regression

The following example shows how to perform each of these types of bivariate analysis in Python using the following pandas DataFrame that contains information about two variables:

(1) Hours spent studying and

(2) Exam score received by 20 different students:

```
import·pandas·as·pd

#create·DataFrame
df·=·pd.DataFrame({'hours':·[1,·1,·1,·2,·2,·2,·3,·3,·3,·3,
·····························3,·4,·4,·5,·5,·6,·6,·6,·7,·8],
···················'score':·[75,·66,·68,·74,·78,·72,·85,·82,·90,·82,
·····························80,·88,·85,·90,·92,·94,·94,·88,·91,·96]})

#view·first·five·rows·of·DataFrame
df.head()
```

	hours	score
0	1	75
1	1	66
2	1	68
3	2	74
4	2	78

```
import pandas as pd

#create DataFrame
df = pd.DataFrame({'hours': [8, 2, 9, 6, 6, 9, 4, 3, 6, 2,
                             9, 8, 4, 5, 5, 6, 6, 6, 7, 8],
                   'score': [75, 66, 68, 74, 78, 72, 85, 82, 90, 82,
                             80, 88, 85, 90, 92, 94, 94, 88, 91, 96]})
```

```
#view first five rows of DataFrame
df.head()
```

	hours	score
0	8	75
1	2	66
2	9	68
3	6	74
4	6	78

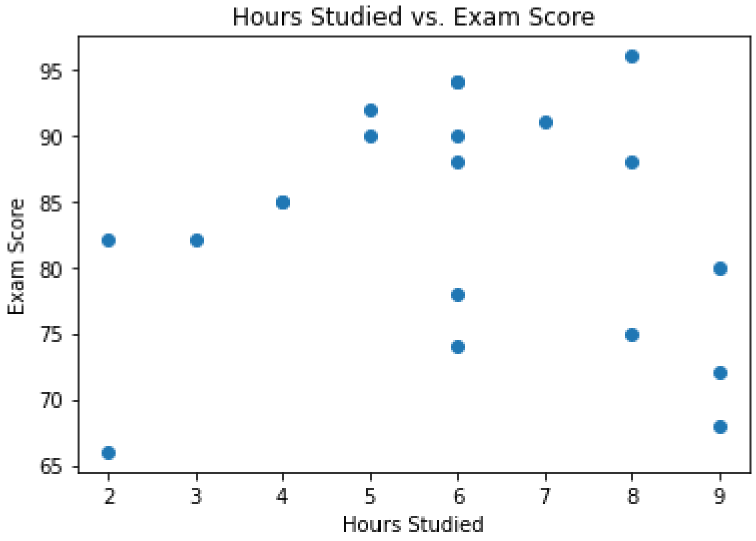
1. Scatterplots

We can use the following syntax to create a scatterplot of hours studied vs. exam score:

```
import·matplotlib.pyplot·as·plt

#create scatterplot of hours vs. score
plt.scatter(df.hours, df.score)
plt.title('Hours Studied vs. Exam Score')
plt.xlabel('Hours Studied')
plt.ylabel('Exam Score')
```

```
Text(0, 0.5, 'Exam Score')
```



The x-axis shows the hours studied and the y-axis shows the exam score received.

From the plot we can see that there is a positive relationship between the two variables: As hours studied increases, exam score tends to increase as well.

2. Correlation Coefficients

A Pearson Correlation Coefficient is a way to quantify the linear relationship between two variables.

We can use the corr() function in pandas to create a correlation matrix:

```
#create correlation matrix
df.corr()
```

	hours	score
hours	1.000000	0.891306
score	0.891306	1.000000

The correlation coefficient turns out to be 0.891.

This indicates a strong positive correlation between hours studied and exam score received.