simple_scatterplot

February 6, 2025

1 Simple Scatterplot

First we must import matplotlib. We do that as follows:

[1]:
```
import matplotlib.pyplot as plt
```

1.1 Sample data

Here we use lists to create some sample data to plot.

[2]:
```
x_values = [1, 2, 3, 4, 5, 6, 10]  # x-axis data
y_values = [2, 3, 5, 7, 11, 13, 14]  # y-axis data
```

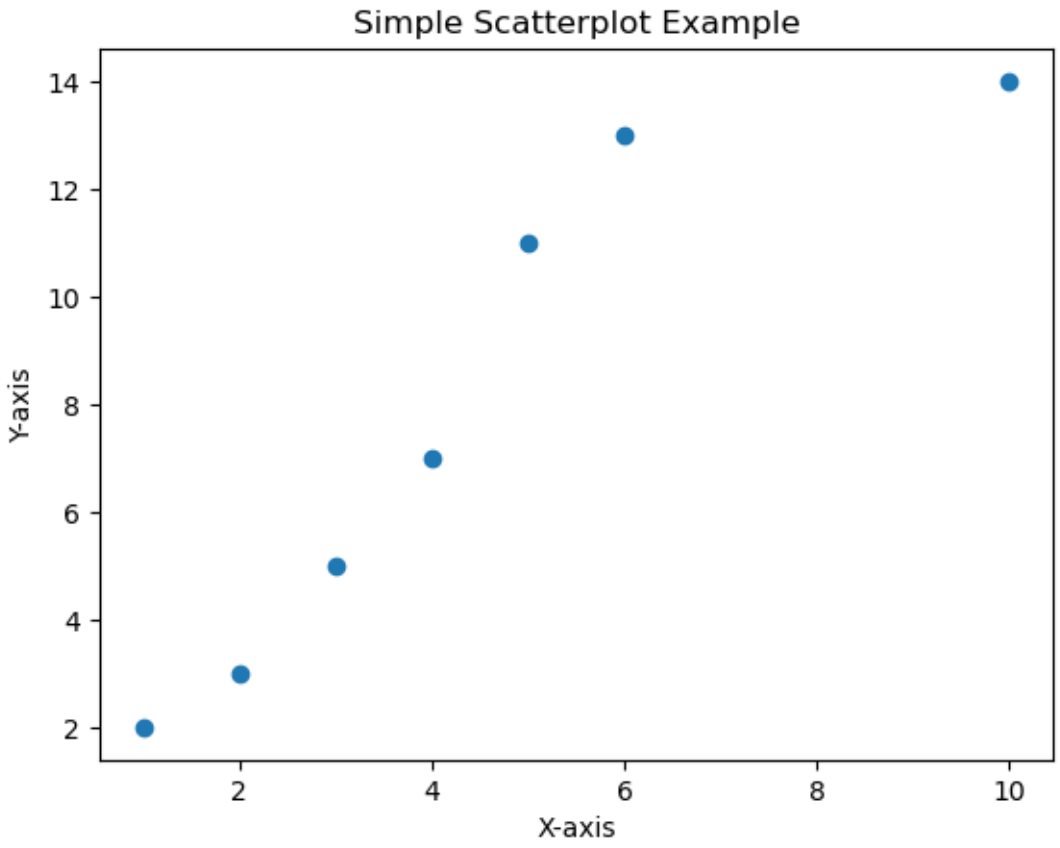
1.2 Plotting the Data

And now we show how to create the scatterplot.

[3]:
```
# Plot the x and y values
plt.scatter(x=x_values,
            y=y_values)

# Adding titles and labels
plt.title("Simple Scatterplot Example")
plt.xlabel("X-axis")
plt.ylabel("Y-axis")

# Displaying the plot
plt.show()
```



1.3 Working with Pandas DataFrames

In the example above we used lists to store our data. Matplotlib is compatible with Pandas as well, so we can plot data stored in a Pandas DataFrame.

[4]:
```
import pandas as pd
```

Here we create a DataFrame called df. Then we create a Series, or column, called 'X' and use our list of x values to add the values. We then add the y values to a 'Y' column.

[5]:
```
df = pd.DataFrame()

df['X'] = x_values
df['Y'] = y_values
```

Let's just demonstrate that we do now have a Pandas DataFrame called df.

[6]:
```
df
```

[6]:

	X	Y
0	1	2

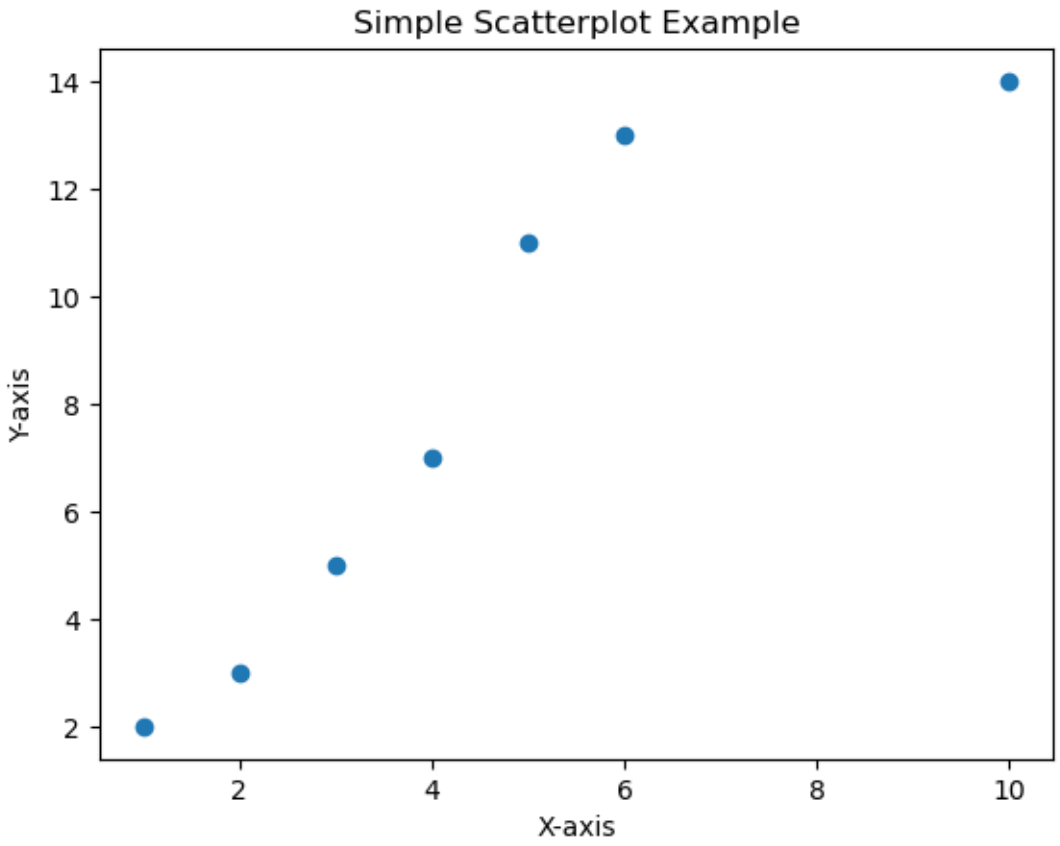
```
1   2   3
2   3   5
3   4   7
4   5  11
5   6  13
6  10  14
```

And now we go ahead and create our Scatterplot from the DataFrame.

[7]:
```
plt.scatter(
    x=df['X'],
    y=df['Y'])

# Adding titles and labels
plt.title("Simple Scatterplot Example")
plt.xlabel("X-axis")
plt.ylabel("Y-axis")

# Displaying the plot
plt.show()
```



1.4 Loading a dataset into a Pandas DataFrame

We might want to plot data from a file or other data source. Here we read a csv file into a Pandas DataFrame.

[8]:
```
sunlight_df = pd.read_csv('sunlight.csv')
```

Examine the first few rows of the DataFrame.

[9]:
```
sunlight_df.head()
```

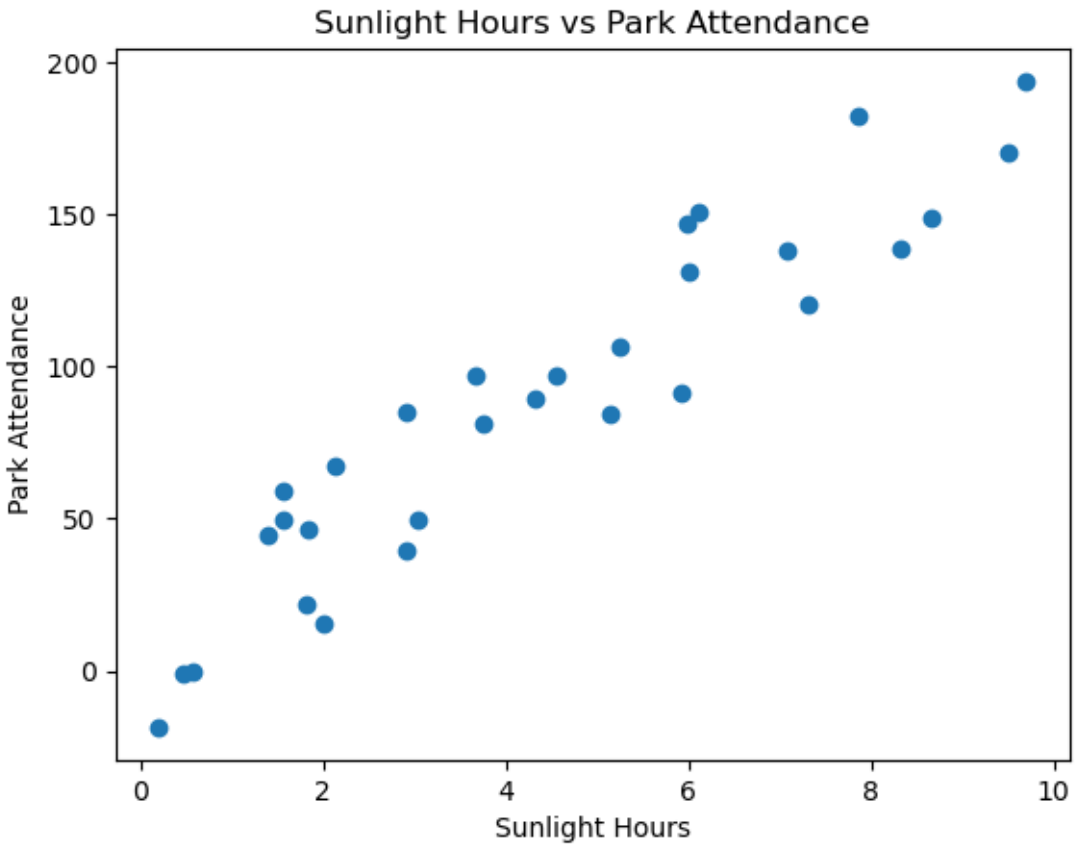
[9]:

	Sunlight Hours	Park Attendance
0	3.745401	81.360715
1	9.507143	170.374309
2	7.319939	120.301884
3	5.986585	146.664829
4	1.560186	59.141650

[10]:
```
plt.scatter(
    x=sunlight_df['Sunlight Hours'],
    y=sunlight_df['Park Attendance'])

# Adding titles and labels
plt.title("Sunlight Hours vs Park Attendance")
plt.xlabel("Sunlight Hours")
plt.ylabel("Park Attendance")

# Displaying the plot
plt.show()
```



Do you notice anything weird about the data plotted above?