In [1]:
```
#Data

x1 = [10.0, 8.0, 13.0, 9.0, 11.0, 14.0, 6.0, 4.0, 12.0, 7.0, 5.0]
y1 = [8.04, 6.95, 7.58, 8.81, 8.33, 9.96, 7.24, 4.26, 10.84, 4.82, 5.6
8]
x2 = [10.0, 8.0, 13.0, 9.0, 11.0, 14.0, 6.0, 4.0, 12.0, 7.0, 5.0]
y2 = [9.14, 8.14, 8.74, 8.77, 9.26, 8.10, 6.13, 3.10, 9.13, 7.26, 4.74
]
x3 = [10.0, 8.0, 13.0, 9.0, 11.0, 14.0, 6.0, 4.0, 12.0, 7.0, 5.0]
y3 = [7.46, 6.77, 12.74, 7.11, 7.81, 8.84, 6.08, 5.39, 8.15, 6.42, 5.7
3]
x4 = [8.0, 8.0, 8.0, 8.0, 8.0, 8.0, 8.0, 19.0, 8.0, 8.0, 8.0]
y4 = [6.58, 5.76, 7.71, 8.84, 8.47, 7.04, 5.25, 12.50, 5.56, 7.91, 6.8
9]
```

# Question 1:

In [2]:
```
import numpy as np
```

In [3]:
```
# Mean
mean_x1 = np.mean(x1)
mean_y1 = np.mean(y1)


mean_x2 = np.mean(x2)
mean_y2 = np.mean(y2)

mean_x3 = np.mean(x4)
mean_y3 = np.mean(y4)

mean_x4 = np.mean(x4)
mean_y4 = np.mean(y4)
```

In [4]:
```
print('Mean: ')
print('x1: ', mean_x1)
print('y1: ', mean_y1)
print('x2: ', mean_x2)
print('y2: ', mean_y2)
print('x3: ', mean_x3)
print('y3: ', mean_y3)
print('x4: ', mean_x4)
print('y4: ', mean_y4)
```

```
Mean:
x1:  9.0
y1:  7.50090909091
x2:  9.0
y2:  7.50090909091
x3:  9.0
y3:  7.50090909091
x4:  9.0
y4:  7.50090909091
```

In [5]:
```
# Variance

var_x1 = np.var(x1)
var_y1 = np.var(y1)

var_x2 = np.var(x2)
var_y2 = np.var(y2)

var_x3 = np.var(x3)
var_y3 = np.var(y3)

var_x4 = np.var(x4)
var_y4 = np.var(y4)
```

In [6]:
```
print('Variance: ')
print('x1: ', var_x1)
print('y1: ', var_y1)
print('x2: ', var_x2)
print('y2: ', var_y2)
print('x3: ', var_x3)
print('y3: ', var_y3)
print('x4: ', var_x4)
print('y4: ', var_y4)
```

```
Variance:
x1:  10.0
y1:  3.75206280992
x2:  10.0
y2:  3.75239008264
x3:  10.0
y3:  3.74783636364
x4:  10.0
y4:  3.74840826446
```

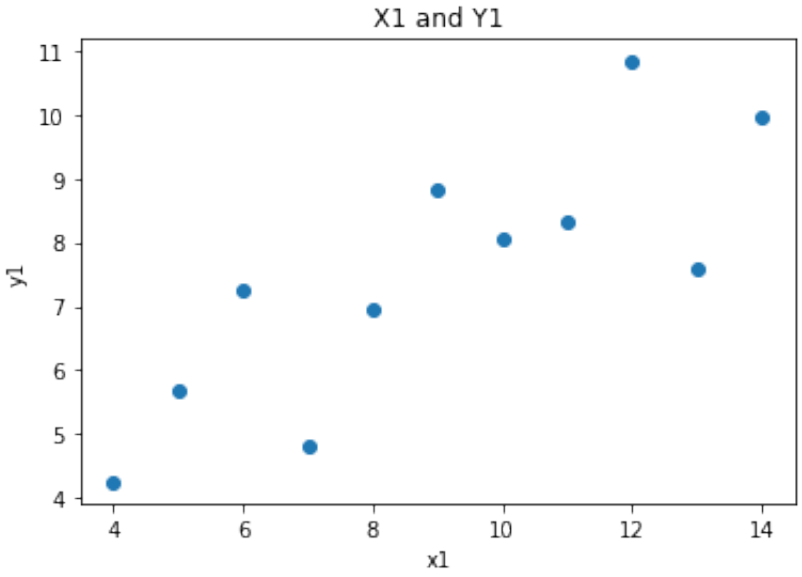
In [7]:
```
# Correlation Coeficient

cor_x1_y1 = np.corrcoef(x1, y1)
cor_x2_y2 = np.corrcoef(x2, y2)
cor_x3_y3 = np.corrcoef(x3, y3)
cor_x4_y4 = np.corrcoef(x4, y4)
```

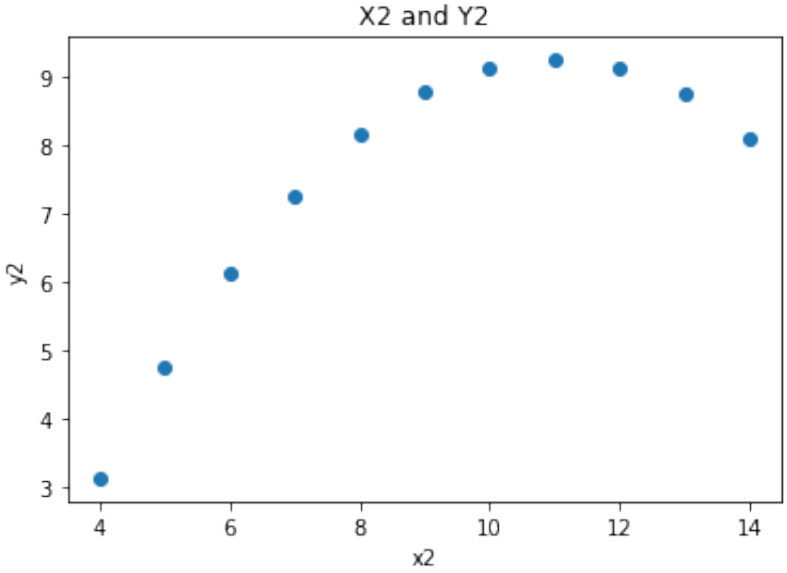
In [8]:
```
print('Correlation: ')
print('x1 and y1: ', cor_x1_y1)
print('x2 and y2: ', cor_x2_y2)
print('x3 and y3: ', cor_x3_y3)
print('x4 and y4: ', cor_x3_y3)
```

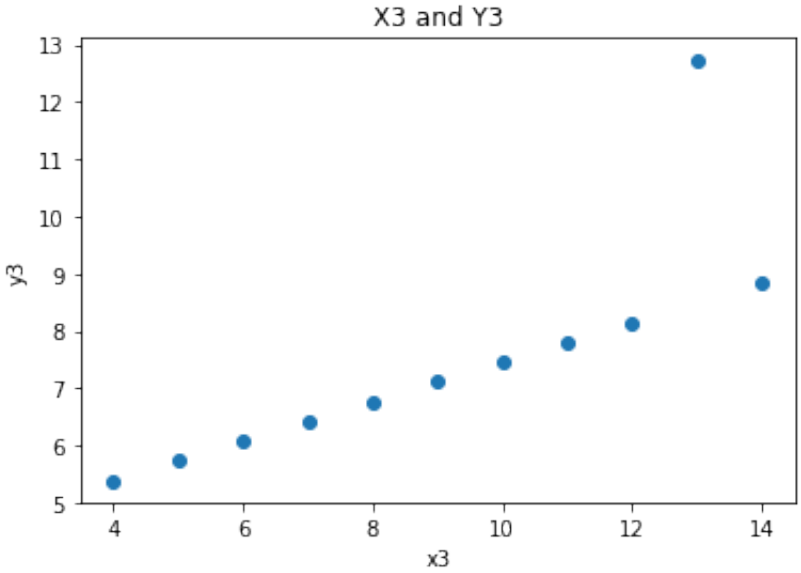
```
Correlation:
x1 and y1:  [[ 1.          0.81642052]
 [ 0.81642052  1.        ]]
x2 and y2:  [[ 1.          0.81623651]
 [ 0.81623651  1.        ]]
x3 and y3:  [[ 1.          0.81628674]
 [ 0.81628674  1.        ]]
x4 and y4:  [[ 1.          0.81628674]
 [ 0.81628674  1.        ]]
```

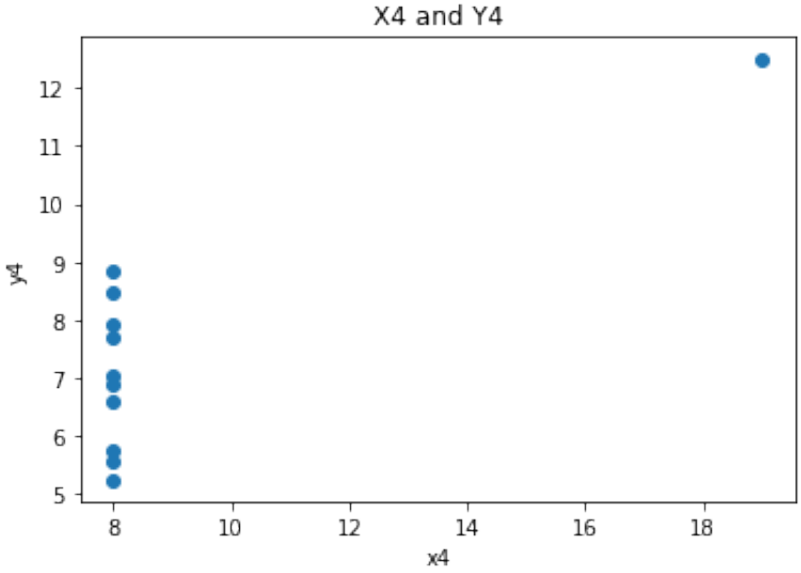
In [9]:
```
import matplotlib.pyplot as plt
```

In [10]:
```
#Scatter for x1 y1
plt.scatter(x1, y1)
plt.title('X1 and Y1')
plt.xlabel('x1')
plt.ylabel('y1')

plt.show()

#Scatter for x2 y2
plt.scatter(x2, y2)
plt.title('X2 and Y2')
plt.xlabel('x2')
plt.ylabel('y2')

plt.show()

#Scatter for x3 y3
plt.scatter(x3, y3)
plt.title('X3 and Y3')
plt.xlabel('x3')
plt.ylabel('y3')

plt.show()

#Scatter for x4 y4
plt.scatter(x4, y4)
plt.title('X4 and Y4')
plt.xlabel('x4')
plt.ylabel('y4')

plt.show()
```




X2 and Y2
y2
x2
3
4
5
6
7
8
9
4
6
8
10
12
14


X3 and Y3
y3
x3
5
6
7
8
9
10
11
12
13
4
6
8
10
12
14


X4 and Y4
y4
x4
5
6
7
8
9
10
11
12
8
10
12
14
16
18

## Question 2:

There is a strong positive linear relationship between each set of data.

## Question 3:

All four sets of data have similar mean, varriance, and coefficent. But those four set data are not identical to each other. There scatter plots are very different.

The first three data sets have the same x variables but different y variables. It seems that the data is trying to find the how different y values relate to certain set of x values. Such as the height of certian specie at certain age (x is age, y is height). The fourth data compare the dfferent y values with one x value. It seems it is exploring how a group of species at certain age have different heights.