

Outliers

Lee el fichero de puntos "outliers.txt" que contiene una serie de puntos en el espacio

Transfórmalo a una numpy array (puedes usar `np.empty` para crear una array vacía y `np.vstack` para ir añadiendo files a una array vacía)

In [92]:

```
[[ 176.  202.]
 [ 220.  120.]
 [ 208.  257.]
 [ 120.  228.]
 [ 155.  177.]
 [ 230.  121.]
 [ 261.  192.]
 [ 151.   94.]
 [ 145.  171.]
 [ 331.  180.]
 [ 229.  149.]
 [ 200.  268.]
 [ 301.  221.]
 [ 239.  129.]
 [ 136.  142.]
 [ 164.  261.]
 [ 223.  218.]
 [ 228.  209.]
 [ 228.  195.]
 [ 184.  171.]
 [ 144.  236.]
 [ 271.  188.]
 [ 272.  162.]
 [ 241.  139.]
 [ 277.  268.]
 [ 153.  310.]
 [ 148.  264.]
 [ 181.  202.]
 [ 249.  130.]
 [ 237.  108.]
 [ 192.   51.]
 [ 110.   80.]
 [  97.  163.]
 [ 139.  203.]
 [ 179.  235.]
 [ 198.  224.]
 [ 227.  159.]
 [ 233.  222.]
 [ 256.  262.]
 [ 273.  262.]
 [ 383.    8.]
 [ 366.  378.]
 [  24.  384.]
 [  27.   32.]]
```

Quédate con la columna de las Xs y de las Ys

In [93]:

```

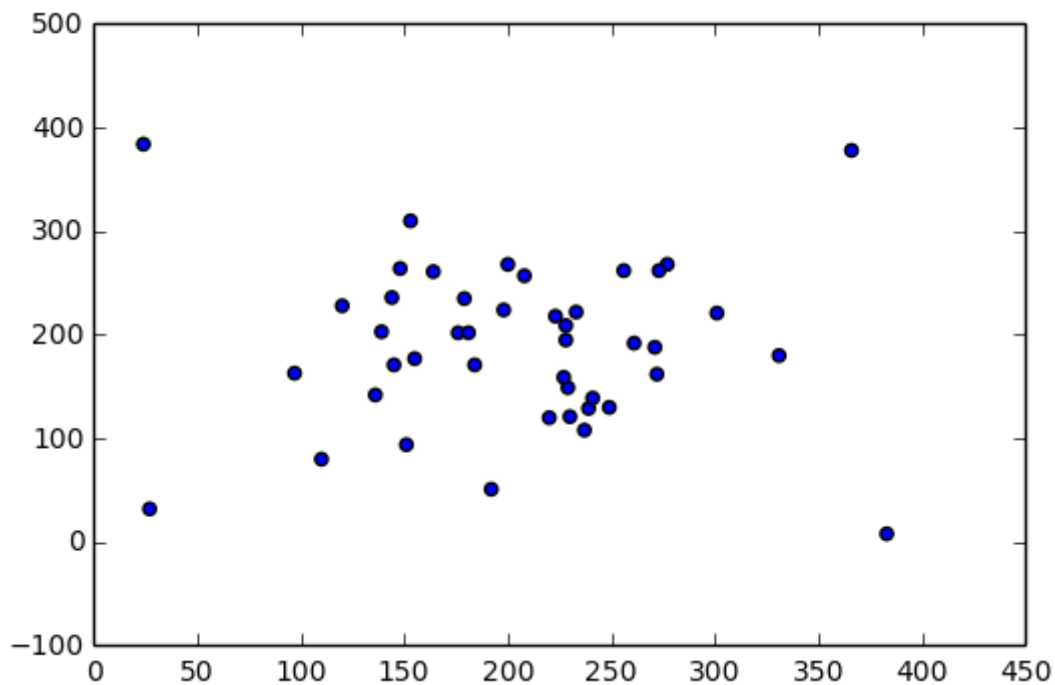
Xs = [ 176.  220.  208.  120.  155.  230.  261.  151.  145.  331.  229.
      200.
      301.  239.  136.  164.  223.  228.  228.  184.  144.  271.  272.  241.
      277.  153.  148.  181.  249.  237.  192.  110.  97.  139.  179.  198.
      227.  233.  256.  273.  383.  366.  24.  27.]
Ys = [ 202.  120.  257.  228.  177.  121.  192.  94.  171.  180.  149.
      268.
      221.  129.  142.  261.  218.  209.  195.  171.  236.  188.  162.  139.
      268.  310.  264.  202.  130.  108.  51.  80.  163.  203.  235.  224.
      159.  222.  262.  262.  8.  378.  384.  32.]

```

Dibuja los valores con una grafica de puntos

In [94]:

Out[94]: <matplotlib.collections.PathCollection at 0x10408fbd0>

**Cálcula el centroide del cluster**

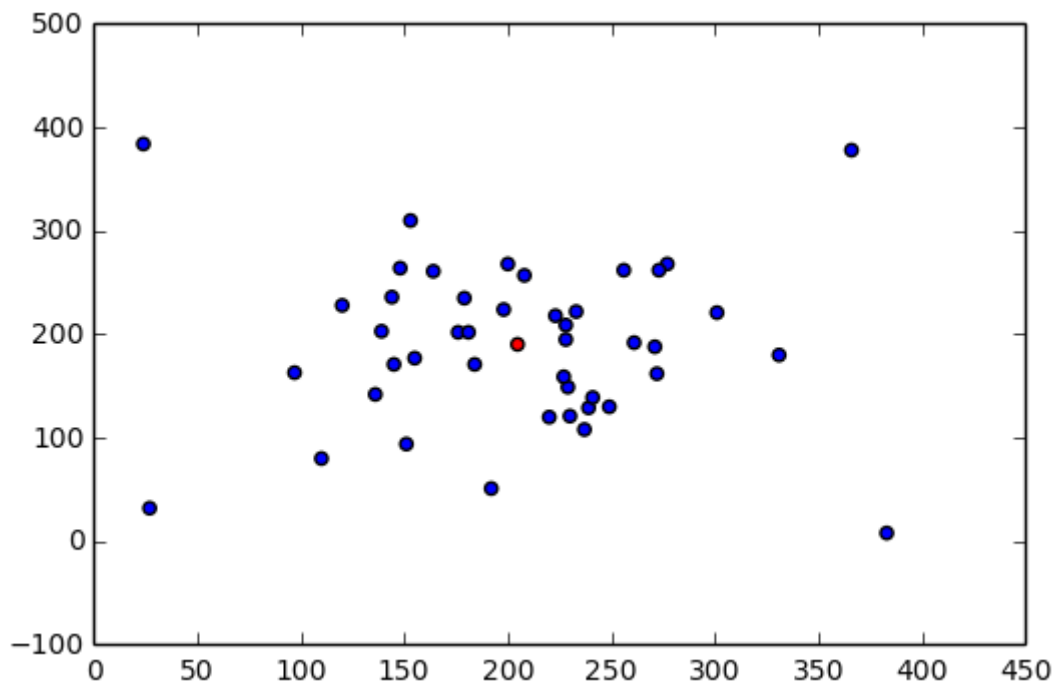
In [95]:

```
[204.68181818181819, 190.34090909090909]
```

Píntalo junto con el resto de los puntos

In [96]:

Out[96]: <matplotlib.collections.PathCollection at 0x104236190>



Cálcula la distancia Euclidiana a de todos los puntos respecto del centroide

In [97]:

In [98]:

```
[ 30.96096082  71.98951442  66.74162668  92.67803115  51.4418401
 73.81850719  56.3426143  110.28738989  62.73747035 126.74074901
 47.96295165  77.80008883 101.08002772  70.28829726  83.98830657
 81.53353579  33.17500708  29.86468277  23.7790818  28.31622093
 75.94100105  66.35948384  73.04070599  62.88798993 106.11717025
130.34304112  92.94347845  26.39626703  74.8673931  88.45614838
139.91682336 145.39519563 111.0986016  66.89061087  51.51689224
 34.31590149  38.47536644  42.47608102  88.13955466  99.0070668
255.0403519  247.46613947 264.85800519 237.99721008]
```

selecciona los puntos que sobrepasen más de un 50% la distancia media al centroide

In [99]:

In [100]:

```
Outlines = [[ 192.  51.]
 [ 110.  80.]
 [ 383.   8.]
 [ 366. 378.]
 [  24. 384.]
 [  27.  32.]]
```

Selecciona los valores buenos, calcula de nuevo el centroid y pintalos

In [101]:

In [102]:

```
[208.0, 195.84210526315789]
```

In [103]:

Out[103]: <matplotlib.collections.PathCollection at 0x1043b20d0>

