In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
```

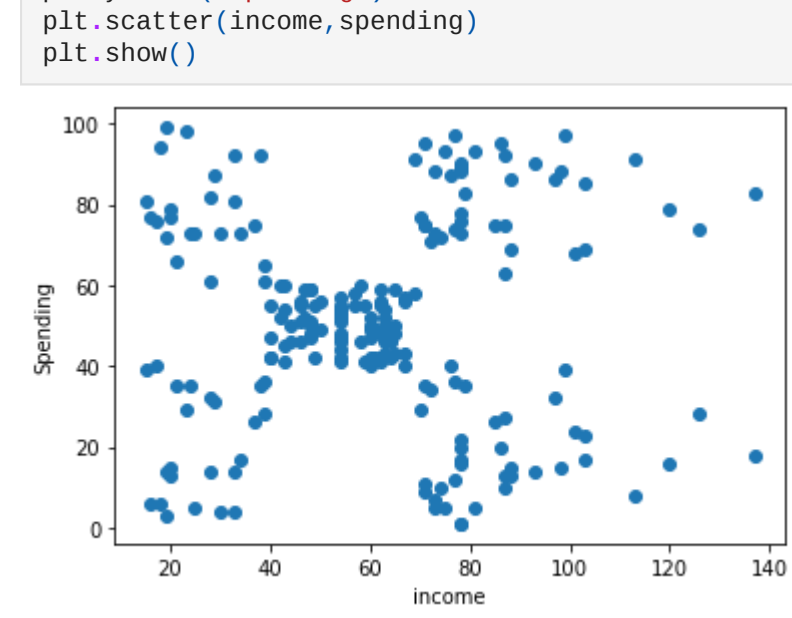
In [2]:
```
data = pd.read_csv("C:\\Users\\Fujitsu-Pc\\OneDrive\\Desktop\\OMEMA FILES\\Spending.csv")
data.head(10)
```

Out[2]:

	CustomerID	Gender	Age	annual_income	SpendingScore	country	region
0	1	Male	19	15	39	Armenia	'Central and Eastern Europe'
1	2	Male	21	15	81	Angola	'Sub-Saharan Africa'
2	3	Female	20	16	6	Argentina	'Latin America and Caribbean'
3	4	Female	23	16	77	Austria	'Western Europe'
4	5	Female	31	17	40	Australia	'Australia and New Zealand'
5	6	Female	22	17	76	Azerbaijan	'Central and Eastern Europe'
6	7	Female	35	18	6	Bangladesh	'Southern Asia'
7	8	Female	23	18	94	Belgium	'Western Europe'
8	9	Male	64	19	3	Burkina Faso	'Sub-Saharan Africa'
9	10	Female	30	19	72	Bulgaria	'Central and Eastern Europe'

In [3]:
```
spending = data['SpendingScore']
income = data['annual_income']
```

In [4]:
```
plt.xlabel('income')
plt.ylabel('Spending')
plt.scatter(income,spending)
plt.show()
```



In [5]:
```
data.sort_values('annual_income',inplace=True)
data.head(6)
```

Out[5]:

	CustomerID	Gender	Age	annual_income	SpendingScore	country	region
0	1	Male	19	15	39	Armenia	'Central and Eastern Europe'
1	2	Male	21	15	81	Angola	'Sub-Saharan Africa'
2	3	Female	20	16	6	Argentina	'Latin America and Caribbean'
3	4	Female	23	16	77	Austria	'Western Europe'
4	5	Female	31	17	40	Australia	'Australia and New Zealand'
5	6	Female	22	17	76	Azerbaijan	'Central and Eastern Europe'

In [6]:
```
richest = data [data['annual_income'] >16]
richest
```

Out[6]:

	CustomerID	Gender	Age	annual_income	SpendingScore	country	region
4	5	Female	31	17	40	Australia	'Australia and New Zealand'
5	6	Female	22	17	76	Azerbaijan	'Central and Eastern Europe'
6	7	Female	35	18	6	Bangladesh	'Southern Asia'
7	8	Female	23	18	94	Belgium	'Western Europe'
11	12	Female	35	19	99	Benin	'Sub-Saharan Africa'
...	...	...	...	...	...	...	...
195	196	Female	35	120	79	Portugal	NaN
196	197	Female	45	126	28	Paraguay	NaN
197	198	Male	32	126	74	Romania	NaN
198	199	Male	32	137	18	Serbia	NaN
199	200	Male	30	137	83	Russia	NaN

196 rows × 7 columns

In [7]:
```
print("THe lowest income in richest: ", richest.iloc[0])
```

```
THe lowest income in richest:  CustomerID                            5
Gender                           Female
Age                                  31
annual_income                        17
SpendingScore                        40
country                       Australia
region           'Australia and New Zealand'
Name: 4, dtype: object
```

In [8]:
```
print("THe highest last country in richest: ", richest.iloc[-1])
```

```
THe highest last country in richest:  CustomerID          200
Gender              Male
Age                   30
annual_income        137
SpendingScore         83
country           Russia
region               NaN
Name: 199, dtype: object
```

In [9]:
```
print("THe highest last country in richest: ", richest.iloc[0:6])
```

```
THe highest last country in richest:       CustomerID  Gender  Age  annual_income  SpendingScore     country  \
4            5  Female   31             17             40   Australia
5            6  Female   22             17             76  Azerbaijan
6            7  Female   35             18              6  Bangladesh
7            8  Female   23             18             94     Belgium
11          12  Female   35             19             99       Benin
10          11    Male   67             19             14     Burundi

                        region
4    'Australia and New Zealand'
5   'Central and Eastern Europe'
6                'Southern Asia'
7               'Western Europe'
11          'Sub-Saharan Africa'
10          'Sub-Saharan Africa'
```

In [10]:
```
print("The mean of richest top countries more than 17K is: ", np.mean(richest['annual_income']))
```

```
The mean of richest top countries more than 17K is:  61.47959183673469
```

In [11]:
```
all_mean = np.mean(data['annual_income'])

all_mean
```

Out[11]: 60.56

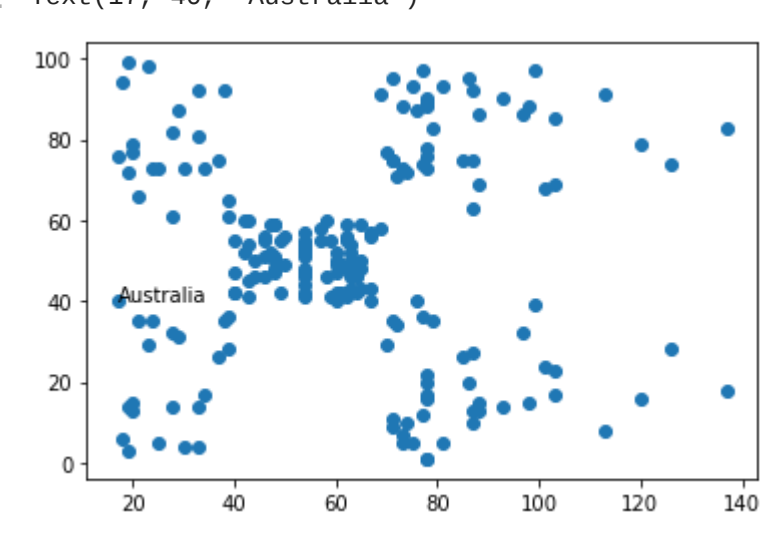
In [12]:
```
print("Both of the means, The mean of top countries: ", np.mean(richest['annual_income']))

print("The mean of All countries: ", np.mean(data['annual_income']))
```

```
Both of the means, The mean of top countries:  61.47959183673469
The mean of All countries:  60.56
```

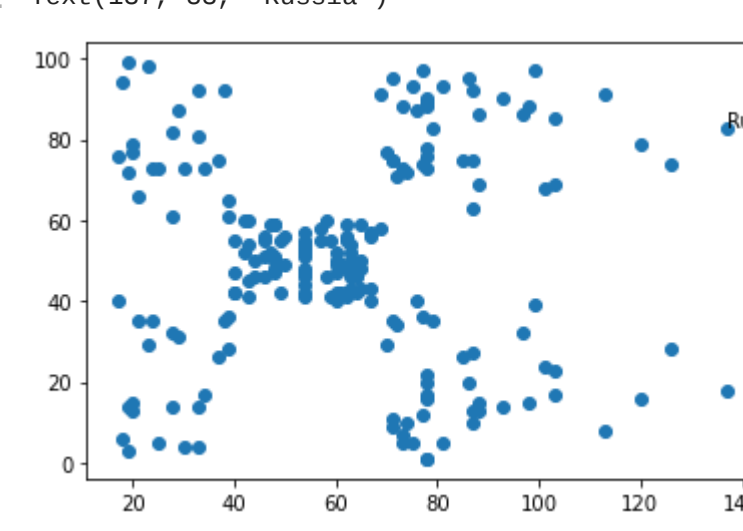
In [13]:
```
plt.scatter(richest['annual_income'],richest['SpendingScore'])
plt.text(richest.iloc[0]['annual_income'],richest.iloc[0]['SpendingScore'],richest.iloc[0]['country'])
```

Out[13]: Text(17, 40, 'Australia')



In [14]:
```
plt.scatter(richest['annual_income'],richest['SpendingScore'])
plt.text(richest.iloc[-1]['annual_income'],richest.iloc[-1]['SpendingScore'],richest.iloc[-1]['country'])
```

Out[14]: Text(137, 83, 'Russia')

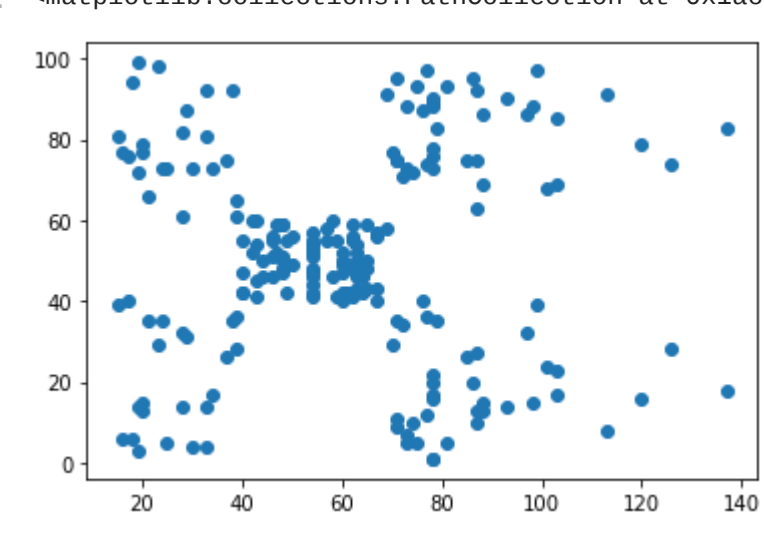


In [15]:
```
for k,row in richest.iterrows():
    print(row['country'])
```

```
Australia
Azerbaijan
Bangladesh
Belgium
Benin
Burundi
Burkina Faso
Bulgaria
Bolivia
Brazil
Botswana
Belarus
Canada
Switzerland
Chile
Cameroon
China
Colombia
Cyprus
Costa Rica
Czech Republic
Germany
Djibouti
Denmark
Dominican Republic
Ecuador
Estonia
Spain
France
Gabon
Ethiopia
Finland
United Kingdom
Georgia
Ghana
Greece
Croatia
Honduras
Ireland
Israel
Hungary
Indonesia
Italy
Iceland
Iran
India
Japan
Kenya
Kyrgyzstan
Cambodia
Kosovo
Kazakhstan
Laos
Sri Lanka
Luxembourg
Lithuania
Liberia
Lesotho
Latvia
Morocco
Moldova
Madagascar
Mali
Mongolia
Mauritania
Mauritius
Malaysia
Mexico
Niger
Mozambique
Nigeria
Paraguay
Portugal
Pakistan
Philippines
Peru
Poland
Nepal
Norway
Netherlands
Nicaragua
Panama
Romania
Serbia
Russia
Rwanda
Sweden
Slovenia
El Salvador
Senegal
Chad
Slovakia
Sierra Leone
Togo
Tajikistan
Thailand
Tunisia
Turkey
Tanzania
Ukraine
Uganda
United States
Zimbabwe
Zambia
Armenia
Vietnam
Uruguay
South Africa
Angola
Argentina
Austria
Australia
Azerbaijan
Bangladesh
Belgium
Burkina Faso
Bulgaria
Burundi
Bolivia
Benin
Botswana
Brazil
Belarus
Canada
Switzerland
Chile
Cameroon
China
Colombia
Costa Rica
Germany
Djibouti
Cyprus
Czech Republic
Denmark
Dominican Republic
Ecuador
Estonia
Spain
Ethiopia
Finland
France
Gabon
United Kingdom
Iran
India
Ghana
Israel
Iceland
Ireland
Georgia
Hungary
Greece
Indonesia
Croatia
Honduras
Italy
Japan
Kenya
Kyrgyzstan
Kosovo
Cambodia
Laos
Kazakhstan
Sri Lanka
Liberia
Lesotho
Lithuania
Luxembourg
Latvia
Mali
Madagascar
Morocco
Moldova
Mongolia
Mauritania
Mauritius
Mexico
Malaysia
Mozambique
Niger
Nigeria
Netherlands
Nicaragua
Norway
Nepal
Panama
Peru
Philippines
Pakistan
Poland
Portugal
Paraguay
Romania
Serbia
Russia
```

In [16]:
```
plt.scatter(income,spending)
```

Out[16]: <matplotlib.collections.PathCollection at 0x1ac9188f580>



In [17]:
```
data = data.drop('CustomerID', axis=1)

from sklearn.preprocessing import LabelEncoder

encode = LabelEncoder()
encoded_sex = encode.fit_transform(data.iloc[:, 0])
print(encoded_sex)
```

```
[1 1 0 0 0 0 0 0 0 1 1 0 0 0 1 1 0 1 1 0 1 1 1 0 0 1 0 1 0 0 1 0 0 0 1 1 0 1 0
 0 0 0 1 0 0 0 1 0 0 0 0 0 1 0 1 0 1 0 1 1 1 0 1 0 0 1 1 0 0 1 0 0 1 0 0 0 0
 1 1 0 1 1 1 0 0 1 0 1 0 0 0 0 0 0 1 1 0 0 1 0 0 1 1 0 0 1 1 1 0 1 1 0 1 0 1 0
 1 0 1 0 0 0 0 0 0 1 0 1 0 0 0 1 1 1 1 1 1 0 0 0 1 1 0 1 0 0 1 0 0 1 1 1 0 1 0
 1 0 1 1 0 0 0 0 1 0 0 1 0 0 1 0 0 1 0 1 0 1 1 1 1 1 1 1 0 0 1 1 0 0 1 0 0 0 0
 1 1 0 0 0 0 0 1 0 0 0 0 1 1 1]
```

In [18]:
```
from sklearn.cluster import KMeans
```

In [19]:
```
income_spending = np.column_stack((income,spending))
print(income_spending)
```

```
[[ 15  39]
 [ 15  81]
 [ 16   6]
 [ 16  77]
 [ 17  40]
 [ 17  76]
 [ 18   6]
 [ 18  94]
 [ 19   3]
 [ 19  72]
 [ 19  14]
 [ 19  99]
 [ 20  15]
 [ 20  77]
 [ 20  13]
 [ 20  79]
 [ 21  35]
 [ 21  66]
 [ 23  29]
 [ 23  98]
 [ 24  35]
 [ 24  73]
 [ 25   5]
 [ 25  73]
 [ 28  14]
 [ 28  82]
 [ 28  32]
 [ 28  61]
 [ 29  31]
 [ 29  87]
 [ 30   4]
 [ 30  73]
 [ 33   4]
 [ 33  92]
 [ 33  14]
 [ 33  81]
 [ 34  17]
 [ 34  73]
 [ 37  26]
 [ 37  75]
 [ 38  35]
 [ 38  92]
 [ 39  36]
 [ 39  61]
 [ 39  28]
 [ 39  65]
 [ 40  55]
 [ 40  47]
 [ 40  42]
 [ 40  42]
 [ 42  52]
 [ 42  60]
 [ 43  54]
 [ 43  60]
 [ 43  45]
 [ 43  41]
 [ 44  50]
 [ 44  46]
 [ 46  51]
 [ 46  46]
 [ 46  56]
 [ 46  55]
 [ 47  52]
 [ 47  59]
 [ 48  51]
 [ 48  59]
 [ 48  50]
 [ 48  48]
 [ 48  59]
 [ 48  47]
 [ 49  55]
 [ 49  42]
 [ 50  49]
 [ 50  56]
 [ 54  47]
 [ 54  54]
 [ 54  53]
 [ 54  48]
 [ 54  52]
 [ 54  42]
 [ 54  51]
 [ 54  55]
 [ 54  41]
 [ 54  44]
 [ 54  57]
 [ 54  46]
 [ 57  58]
 [ 57  55]
 [ 58  60]
 [ 58  46]
 [ 59  55]
 [ 59  41]
 [ 60  49]
 [ 60  40]
 [ 60  42]
 [ 60  52]
 [ 60  47]
 [ 60  50]
 [ 61  42]
 [ 61  49]
 [ 62  41]
 [ 62  48]
 [ 62  59]
 [ 62  55]
 [ 62  56]
 [ 62  42]
 [ 63  50]
 [ 63  46]
 [ 63  43]
 [ 63  48]
 [ 63  52]
 [ 63  54]
 [ 64  42]
 [ 64  46]
 [ 65  48]
 [ 65  50]
 [ 65  43]
 [ 65  59]
 [ 67  43]
 [ 67  57]
 [ 67  56]
 [ 67  40]
 [ 69  58]
 [ 69  91]
 [ 70  29]
 [ 70  77]
 [ 71  35]
 [ 71  95]
 [ 71  11]
 [ 71  75]
 [ 71   9]
 [ 71  75]
 [ 72  34]
 [ 72  71]
 [ 73   5]
 [ 73  88]
 [ 73   7]
 [ 73  73]
 [ 74  10]
 [ 74  72]
 [ 75   5]
 [ 75  93]
 [ 76  40]
 [ 76  87]
 [ 77  12]
 [ 77  97]
 [ 77  36]
 [ 77  74]
 [ 78  22]
 [ 78  90]
 [ 78  17]
 [ 78  88]
 [ 78  20]
 [ 78  76]
 [ 78  16]
 [ 78  89]
 [ 78   1]
 [ 78  78]
 [ 78   1]
 [ 78  73]
 [ 79  35]
 [ 79  83]
 [ 81   5]
 [ 81  93]
 [ 85  26]
 [ 85  75]
 [ 86  20]
 [ 86  95]
 [ 87  27]
 [ 87  63]
 [ 87  13]
 [ 87  75]
 [ 87  10]
 [ 87  92]
 [ 88  13]
 [ 88  86]
 [ 88  15]
 [ 88  69]
 [ 93  14]
 [ 93  90]
 [ 97  32]
 [ 97  86]
 [ 98  15]
 [ 98  88]
 [ 99  39]
 [ 99  97]
 [101  24]
 [101  68]
 [103  17]
 [103  85]
 [103  23]
 [103  69]
 [113   8]
 [113  91]
 [120  16]
 [120  79]
 [126  28]
 [126  74]
 [137  18]
 [137  83]]
```

In [21]:
```
km_results = KMeans(n_clusters=2).fit(income_spending)
km_results
```

Out[21]: KMeans(n_clusters=2)

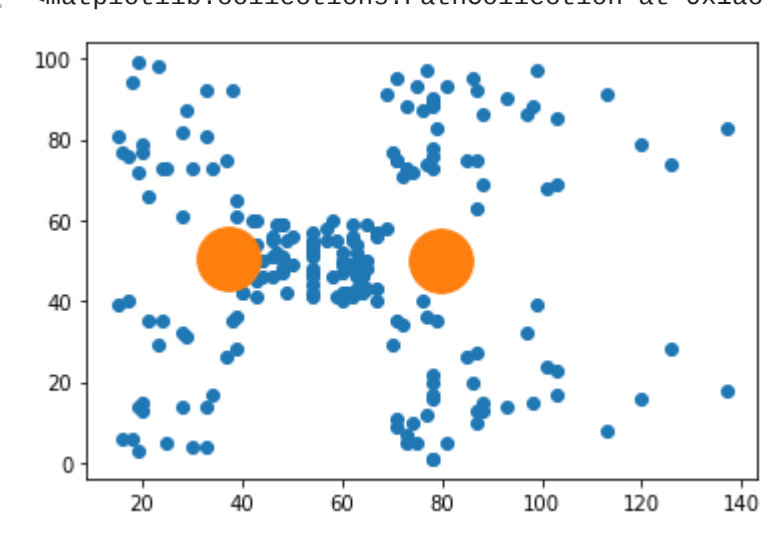
In [22]:
```
km_results.cluster_centers_
```

Out[22]:
```
array([[79.6       , 50.12727273],
       [37.28888889, 50.28888889]])
```

In [24]:
```
clusters = km_results.cluster_centers_

plt.scatter(income,spending)
plt.scatter(clusters[:,0],clusters[:,1],s=1000)
```

Out[24]: <matplotlib.collections.PathCollection at 0x1ac944842b0>



In [ ]:

In [ ]: