

```
In [61]: import numpy as np
import pandas
from pandas import DataFrame, Series
import statsmodels.formula.api as sm
from sklearn.linear_model import LinearRegression
import scipy, scipy.stats
#import matplotlib.pyplot as plt
%matplotlib inline
```

```
In [62]: data_str = '''Region Alcohol Tobacco
North 6.47 4.03
Yorkshire 6.13 3.76
Northeast 6.19 3.77
East-Midlands 4.89 3.34
West-Midlands 5.63 3.47
East-Anglia 4.52 2.92
Southeast 5.89 3.20
Southwest 4.79 2.71
Wales 5.27 3.53
Scotland 6.08 4.51
Northern-Ireland 4.02 4.56'''
```

```
In [63]: d = data_str.split('\n')
```

```
In [64]: d
```

```
Out[64]: ['Region Alcohol Tobacco',
'North 6.47 4.03',
'Yorkshire 6.13 3.76',
'Northeast 6.19 3.77',
'East-Midlands 4.89 3.34',
'West-Midlands 5.63 3.47',
'East-Anglia 4.52 2.92',
'Southeast 5.89 3.20',
'Southwest 4.79 2.71',
'Wales 5.27 3.53',
'Scotland 6.08 4.51',
'Northern-Ireland 4.02 4.56']
```

```
In [65]: d = [ i.split(' ') for i in d ]
```

In [66]: d

```
Out[66]: [['Region', 'Alcohol', 'Tobacco'],
          ['North', '6.47', '4.03'],
          ['Yorkshire', '6.13', '3.76'],
          ['Northeast', '6.19', '3.77'],
          ['East-Midlands', '4.89', '3.34'],
          ['West-Midlands', '5.63', '3.47'],
          ['East-Anglia', '4.52', '2.92'],
          ['Southeast', '5.89', '3.20'],
          ['Southwest', '4.79', '2.71'],

          ['Wales', '5.27', '3.53'],
          ['Scotland', '6.08', '4.51'],
          ['Northern-Ireland', '4.02', '4.56']]
```

```
In [67]: for i in range( len(d)):
          for j in range(len(d[0])):

              #print j
              try:
                  d[i][j] = float(d[i][j])
              except:
                  pass
```

```
In [68]: df = DataFrame ( d[1:], columns =d[0])
```

In [69]: df

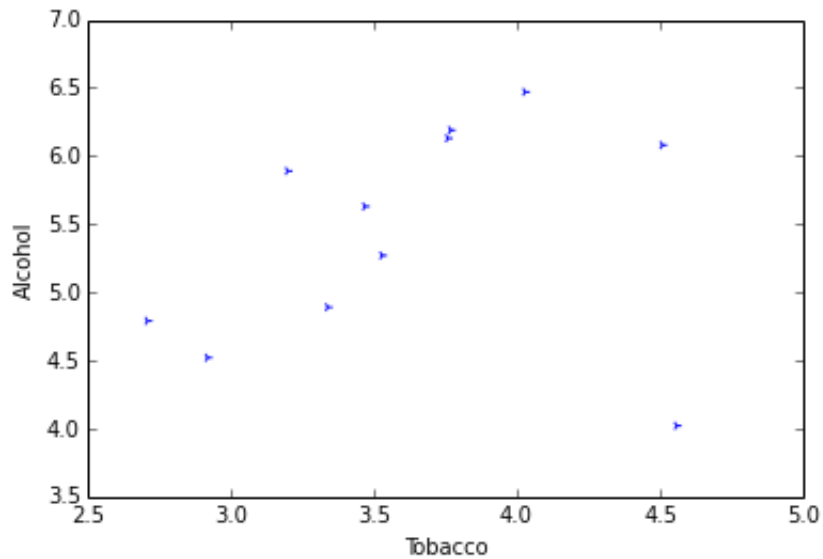
```
Out[69]:
```

	Region	Alcohol	Tobacco
0	North	6.47	4.03
1	Yorkshire	6.13	3.76
2	Northeast	6.19	3.77
3	East-Midlands	4.89	3.34
4	West-Midlands	5.63	3.47
5	East-Anglia	4.52	2.92
6	Southeast	5.89	3.20
7	Southwest	4.79	2.71
8	Wales	5.27	3.53
9	Scotland	6.08	4.51
10	Northern-Ireland	4.02	4.56

```
In [70]: scatter( df.Tobacco, df.Alcohol,
                  marker='2',
                  edgecolor='b',
                  facecolor='none', alpha = 0.9)

xlabel('Tobacco')
ylabel('Alcohol')
#savefig('alcohol_v_tobacco.png', fmt='png', dpi=100)
```

Out[70]: <matplotlib.text.Text at 0x7f1437e5f450>



In []: