Lesson 6

Lets take a look at the ***groupby*** function.

In [1]:
```
# Import libraries
import pandas as pd
import sys
```

In [2]:
```
print 'Python version ' + sys.version
print 'Pandas version: ' + pd.__version__
```

```
Python version 2.7.5 |Anaconda 2.1.0 (64-bit)| (default, Jul  1 2013, 1
Pandas version: 0.15.2
```

In [3]:
```
# Our small data set
d = {'one':[1,1,1,1,1],
     'two':[2,2,2,2,2],
     'letter':['a','a','b','b','c']}

# Create dataframe
df = pd.DataFrame(d)
df
```

Out[3]:

	letter	one	two
0	a	1	2
1	a	1	2
2	b	1	2
3	b	1	2
4	c	1	2

In [4]:
```
# Create group object
one = df.groupby('letter')

# Apply sum function
one.sum()
```

Out[4]:

	one	two
letter		
a	2	4
b	2	4
c	1	2

In [5]:
```
letterone = df.groupby(['letter','one']).sum()
letterone
```

Out[5]:

		two
letter	**one**	
a	**1**	4
b	**1**	4
c	**1**	2

In [6]:
```
letterone.index
```

Out[6]:
```
MultiIndex(levels=[[u'a', u'b', u'c'], [1]],
           labels=[[0, 1, 2], [0, 0, 0]],
           names=[u'letter', u'one'])
```

You may want to ***not*** have the columns you are grouping by become your index, this can be easily achieved as shown below.

In [7]:
```
letterone = df.groupby(['letter','one'], as_index=False).sum()
letterone
```

Out[7]:

	letter	one	two
0	a	1	4
1	b	1	4
2	c	1	2

In [8]:
```
letterone.index
```

Out[8]:
```
Int64Index([0, 1, 2], dtype='int64')
```

Author: David Rojas (http://www.hedaro.com/)