CREATING A DATAFRAME WITH MULTI-LEVEL INDEXES

In [23]:
```
import numpy as np
import pandas as pd
from numpy.random import randn
```

In [51]:
```
outside_index = 'SIM1 SIM1 SIM1 SIM1 SIM2 SIM2 SIM2 SIM2'.split()
```

In [52]:
```
outside_index
```

Out[52]:
```
['SIM1', 'SIM1', 'SIM1', 'SIM1', 'SIM2', 'SIM2', 'SIM2', 'SIM2']
```

In [53]:
```
inside_index = [1,2,3,4,1,2,3,4]
```

In [54]:
```
inside_index
```

Out[54]:
```
[1, 2, 3, 4, 1, 2, 3, 4]
```

In [55]:
```
#creating a list of tuple pairs
```

In [56]:
```
hier_index = list(zip(outside_index, inside_index))
```

In [57]:
```
hier_index = pd.MultiIndex.from_tuples(hier_index)
```

In [58]:
```
#creating the dataframe
```

In [59]:
```
df = pd.DataFrame(randn(8,3),hier_index,['Score1', 'Score2', 'Score3'])
```

In [60]:
```
df
```

Out[60]:

		Score1	Score2	Score3
SIM1	**1**	-0.232674	-0.522676	1.548499
	2	-1.532651	1.708316	0.005002
	3	-0.671043	2.360634	-1.127717
	4	2.117229	0.982231	-1.485692
SIM2	**1**	0.188956	-0.871423	0.680666
	2	-0.169509	-0.706081	-0.167604
	3	0.161521	-0.111672	0.008180
	4	0.236829	-0.562505	0.690904

In [63]:
```
#Call simulation1
df.loc['SIM1']
```

Out[63]:

	Score1	Score2	Score3
1	-0.232674	-0.522676	1.548499
2	-1.532651	1.708316	0.005002
3	-0.671043	2.360634	-1.127717
4	2.117229	0.982231	-1.485692

In [66]:
```
#Call trial 3 from simulation2
df.loc['SIM2'].loc[3]
```

Out[66]:
```
Score1    0.161521
Score2   -0.111672
Score3    0.008180
Name: 3, dtype: float64
```

In [67]:
```
#to name the indexes
df.index.names = ['Sim Group', 'Trial']
```

In [68]:
```
#call dataframe again
df
```

Out[68]:

		Score1	Score2	Score3
Sim Group	**Trial**			
SIM1	**1**	-0.232674	-0.522676	1.548499
	2	-1.532651	1.708316	0.005002
	3	-0.671043	2.360634	-1.127717
	4	2.117229	0.982231	-1.485692
SIM2	**1**	0.188956	-0.871423	0.680666
	2	-0.169509	-0.706081	-0.167604
	3	0.161521	-0.111672	0.008180
	4	0.236829	-0.562505	0.690904

In [70]:
```
#cross-sections can skip inside a multi-level index
#the following will return trial 1 from both SIM1 and SIM2
df.xs(1, level = 'Trial')
```

Out[70]:

	Score1	Score2	Score3
Sim Group			
SIM1	-0.232674	-0.522676	1.548499
SIM2	0.188956	-0.871423	0.680666

Prepared by Gary-Gregoire Coquillo

If you want to use Data Science to enhance your business let’s connect and learn together!