Hierarchical Indexing

August 21, 2022

[3]:
```
import pandas as pd
import numpy as np
data = pd.Series(np.random.randn(9),index=[['a', 'a', 'a', 'b', 'b', 'c', 'c',␣
↪'d', 'd'],
                                           [1, 2, 3, 1, 3, 1, 2, 2, 3]])
```

[4]:
```
data
```

[4]:
```
a  1   -1.703189
   2   -1.359302
   3   -0.887155
b  1   -1.040745
   3    0.661998
c  1   -0.188556
   2    0.009014
d  2   -0.913749
   3   -0.786489
dtype: float64
```

[5]:
```
data.index
```

[5]:
```
MultiIndex([('a', 1),
            ('a', 2),
            ('a', 3),
            ('b', 1),
            ('b', 3),
            ('c', 1),
            ('c', 2),
            ('d', 2),
            ('d', 3)],
           )
```

[6]:
```
data['b']
```

[6]:
```
1   -1.040745
3    0.661998
dtype: float64
```

[8]:
```
data['b':'c']
```

[8]:
```
b  1   -1.040745
   3    0.661998
c  1   -0.188556
   2    0.009014
dtype: float64
```

[15]:
```
data.loc[['b','d']]
```

[15]:
```
b  1   -1.040745
   3    0.661998
d  2   -0.913749
   3   -0.786489
dtype: float64
```

[14]:
```
data[:,2]
```

[14]:
```
a   -1.359302
c    0.009014
d   -0.913749
dtype: float64
```

[16]:
```
frame = pd.DataFrame(np.arange(12).reshape((4, 3)),
                     index=[['a', 'a', 'b', 'b'], [1, 2, 1, 2]],
                      columns=[['Ohio', 'Ohio', 'Colorado'],
                      ['Green', 'Red', 'Green']])
```

[17]:
```
frame
```

[17]:
```
      Ohio     Colorado
     Green Red    Green
a 1      0   1        2
  2      3   4        5
b 1      6   7        8
  2      9  10       11
```

[22]:
```
frame.index.names = ['key1','key2']
frame.columns.names = ['state','color']
```

[23]:
```
frame
```

[23]:
```
state      Ohio     Colorado
color     Green Red    Green
key1 key2
a    1        0   1        2
     2        3   4        5
```

```
b    1        6   7        8
     2        9  10       11
```

[25]:
```
frame['Ohio']
```

[25]:
```
color      Green  Red
key1 key2
a    1         0    1
     2         3    4
b    1         6    7
     2         9   10
```

[26]:
```
frame.swaplevel('key1', 'key2')
```

[26]:
```
state      Ohio     Colorado
color     Green Red    Green
key2 key1
1    a        0   1        2
2    a        3   4        5
1    b        6   7        8
2    b        9  10       11
```

[27]:
```
frame.sort_index(level=1)
```

[27]:
```
state      Ohio     Colorado
color     Green Red    Green
key1 key2
a    1        0   1        2
b    1        6   7        8
a    2        3   4        5
b    2        9  10       11
```

[28]:
```
 frame.swaplevel(0, 1).sort_index(level=0)
```

[28]:
```
state      Ohio     Colorado
color     Green Red    Green
key2 key1
1    a        0   1        2
     b        6   7        8
2    a        3   4        5
     b        9  10       11
```

[29]:
```
frame = pd.DataFrame({'a': range(7), 'b': range(7, 0, -1),
                      'c': ['one', 'one', 'one', 'two', 'two',
                            'two', 'two'], 'd': [0, 1, 2, 0, 1, 2, 3]})
```

[30]:
```
frame
```

[30]:
```
   a  b    c  d
0  0  7  one  0
1  1  6  one  1
2  2  5  one  2
3  3  4  two  0
4  4  3  two  1
5  5  2  two  2
6  6  1  two  3
```

[31]:
```
frame2 = frame.set_index(['c', 'd'])
```

[32]:
```
frame2
```

[32]:
```
       a  b
c   d
one 0  0  7
    1  1  6
    2  2  5
two 0  3  4
    1  4  3
    2  5  2
    3  6  1
```

[33]:
```
frame.set_index(['c', 'd'], drop=False)
```

[33]:
```
       a  b    c  d
c   d
one 0  0  7  one  0
    1  1  6  one  1
    2  2  5  one  2
two 0  3  4  two  0
    1  4  3  two  1
    2  5  2  two  2
    3  6  1  two  3
```

[34]:
```
df1 = pd.DataFrame({'key': ['b', 'b', 'a', 'c', 'a', 'a', 'b'],
                    'data1': range(7)})
```

[35]:
```
df2 = pd.DataFrame({'key': ['a', 'b', 'd'],
                    'data2': range(3)})
```

[36]:
```
df1
```

[36]:
```
  key  data1
0   b      0
1   b      1
2   a      2
```

```
3   c      3
4   a      4
5   a      5
6   b      6
```

[37]:
```
df2
```

[37]:
```
  key  data2
0   a      0
1   b      1
2   d      2
```

[42]:
```
pd.merge(df1,df2)
```

[42]:
```
  key  data1  data2
0   b      0      1
1   b      1      1
2   b      6      1
3   a      2      0
4   a      4      0
5   a      5      0
```

[40]:
```
c.sort_index(level =0)
```

[40]:
```
  key  data1  data2
0   b      0      1
1   b      1      1
2   b      6      1
3   a      2      0
4   a      4      0
5   a      5      0
```

[44]:
```
pd.merge(df1, df2, how='right')
```

[44]:
```
  key  data1  data2
0   a    2.0      0
1   a    4.0      0
2   a    5.0      0
3   b    0.0      1
4   b    1.0      1
5   b    6.0      1
6   d    NaN      2
```

[46]:
```
left = pd.DataFrame({'key1': ['foo', 'foo', 'bar'],
                     'key2': ['one', 'two', 'one'],
                        'lval': [1, 2, 3]})
```

[47]:
```
right = pd.DataFrame({'key1': ['foo', 'foo', 'bar', 'bar'],
                      'key2': ['one', 'one', 'one', 'two'],
                      'rval': [4, 5, 6, 7]})
```

[48]:
```
pd.merge(left, right, on=['key1', 'key2'], how='outer')
```

[48]:

	key1	key2	lval	rval
0	foo	one	1.0	4.0
1	foo	one	1.0	5.0
2	foo	two	2.0	NaN
3	bar	one	3.0	6.0
4	bar	two	NaN	7.0

[49]:
```
left
```

[49]:

	key1	key2	lval
0	foo	one	1
1	foo	two	2
2	bar	one	3

[50]:
```
right
```

[50]:

	key1	key2	rval
0	foo	one	4
1	foo	one	5
2	bar	one	6
3	bar	two	7

[51]:
```
pd.merge(df1, df2, on='key', how='left')
```

[51]:

	key	data1	data2
0	b	0	1.0
1	b	1	1.0
2	a	2	0.0
3	c	3	NaN
4	a	4	0.0
5	a	5	0.0
6	b	6	1.0

[ ]:
```

```