Pandas Basics

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```
s = pd.Series(np.random.randn(5), index=['a','b', 'c', 'd', 'e'])
```

```
s
```

```
a    0.862515
b    0.744877
c   -1.231010
d    1.029128
e    0.698776
dtype: float64
```

```
s1 = pd.Series(np.random.randn(5))
s1
```

```
0    0.294792
1    0.296205
2   -0.905825
3   -1.220514
4   -0.957423
dtype: float64
```

```
s
```

```
a    0.485907
b    0.380973
c    1.505020
d    0.651924
e    0.394247
dtype: float64
```

```
s.index
```

```
Index(['a', 'b', 'c', 'd', 'e'], dtype='object')
```

```
pd.Series(np.random.randn(5))
```

```
0    0.043536
1   -2.289564
```

```
d = {'a' : 0., 'b' : 1., 'c' : 2.}
```

```
4    0.754432
```

```
pd.Series(d)
```

```
a    0.0
b    1.0
c    2.0
dtype: float64
```

```
pd.Series(d, index=['b', 'c', 'd', 'a'])
```

```
b    1.0
c    2.0
d    NaN
a    0.0
dtype: float64
```

```
pd.Series(5., index=['a', 'b', 'c', 'd', 'e'])
```

```
a    5.0
b    5.0
c    5.0
d    5.0
e    5.0
dtype: float64
```

```
s[0]
```

```
0.4859073134858728
```

```
s[:3]
```

```
a    0.485907
b    0.380973
c    1.505020
dtype: float64
```

```
s['a']
```

```
0.4859073134858728
```

```
s['e'] = 12.
```

```
s
```

```
a     0.485907
b     0.380973
c     1.505020
d     0.651924
e    12.000000
dtype: float64
```

```
s.get('a')
```

```
0.4859073134858728
```

+ Code + Text

```
ts1 = pd.Series(np.random.randn(5))
ts2 = pd.Series(np.random.randn(5))
```

```
ts1
```

```
0    1.423382
1    0.853949
2   -0.919942
3    1.085223
4   -0.335189
dtype: float64
```

```
d = {'col1': ts1, 'col2': ts2}
d
```

```
{'col1': 0    1.423382
 1    0.853949
 2   -0.919942
 3    1.085223
 4   -0.335189
 dtype: float64,
 'col2': 0    1.652744
 1   -0.263264
 2   -0.055775
 3    0.463031
 4    1.590819
 dtype: float64}
```

```
df1 = pd.DataFrame(data = d)
df1
```

```
df2 = pd.DataFrame(np.random.randn(10, 5))
df2
```

	0	1	2	3	4
0	-0.381261	0.033780	0.528511	0.117758	-0.512066
1	0.800685	-1.171704	-0.191590	1.617451	-0.541814
2	-1.373350	-0.743076	1.593876	0.384355	-1.112257
3	-0.898219	1.116280	-0.024264	-0.369921	-0.770192
4	0.893150	0.309397	1.659716	1.486194	1.313696
5	-0.943646	1.382572	1.336313	1.537604	-0.347292
6	0.511344	0.518097	-0.066484	0.040798	1.135914
7	-0.708785	-0.821855	0.040576	0.799110	2.320810
8	-0.912141	1.734592	-0.186494	0.375898	1.225296
9	0.258302	0.756404	0.371488	0.322876	1.337919

```
df3 = pd.DataFrame(np.random.randn(10, 5), columns=['a', 'b', 'c', 'd', 'e'])
df3
```

	a	b	c	d	e
0	-0.647493	-0.381582	-0.319376	-0.901886	-0.992509
1	1.861222	-0.492064	-0.117981	-0.294205	-0.197586
2	-1.293241	-0.096428	0.553612	0.548095	0.256423
3	0.647029	-0.692647	-0.438434	0.274016	-2.107366
4	0.879660	1.349460	-1.017092	-0.396708	0.510180
5	-0.570104	0.208099	-0.977652	-2.000521	-0.176432
6	-1.130892	-0.270368	1.154039	-0.703176	-1.335627
7	0.490532	2.624482	-0.007277	0.500553	0.487479
8	-0.589384	0.013179	-1.061376	-1.275024	1.869213
9	-0.823418	-0.313433	2.030221	-1.152295	-0.535012

```
d = {'one' : pd.Series([1., 2., 3.], index=['a', 'b', 'c']), 'two' : pd.Series([1., 2
df = pd.DataFrame(d)
df
```

	one	two
a	1.0	1.0
b	2.0	2.0
c	3.0	3.0
d	NaN	4.0

```
pd.DataFrame(d, index=['d', 'b', 'a'])
```

	one	two
d	NaN	4.0
b	2.0	2.0
a	1.0	1.0

```
df.index
```

```
Index(['a', 'b', 'c', 'd'], dtype='object')
```

```
df.columns
```

```
Index(['one', 'two'], dtype='object')
```

```
df.index.hasnans
```

```
False
```

```
dfc = pd.read_csv('data1.csv')
dfc
```

	date	variable	value
0	2000-01-03	A	0.469112
1	2000-01-04	A	-0.282863
2	2000-01-05	A	-1.509059

```
dfc['variable'] == 'A'
```

```
0      True
1      True
2      True
3     False
4     False
5     False
6     False
7     False
8     False
9     False
10    False
11    False
Name: variable, dtype: bool
```

```
dfc[dfc['variable'] == 'A']
```

	date	variable	value
0	2000-01-03	A	0.469112
1	2000-01-04	A	-0.282863
2	2000-01-05	A	-1.509059

```
dfc.pivot(index='date', columns='variable', values='value')
```

variable	A	B	C	D
date				
2000-01-03	0.469112	-1.135632	0.119209	-2.104569
2000-01-04	-0.282863	1.212112	-1.044236	-0.494929
2000-01-05	-1.509059	-0.173215	-0.861849	1.071804

```
dfc.describe()
```

	value
count	12.000000
mean	-0.394510
std	1.007649
min	-2.104569
25%	-1.067085
50%	-0.388896
75%	0.206685
max	1.212112