

Creating Toy Data

November 20, 2016

```
In [1]: import pandas as pd
```

```
In [5]: # key is column name and value is series of values
df = pd.DataFrame({'id' : [100, 101, 102], 'color': ['red', 'blue', 'red']})
```

```
In [6]: df
```

```
Out[6]:
```

	id	color
a	100	red
b	101	blue
c	102	red

```
In [8]: pd.DataFrame([[100, 'red'], [101, 'blue'], [102, 'red']], columns = ['id', 'color'])
```

```
Out[8]:
```

	id	color
0	100	red
1	101	blue
2	102	red

```
In [9]: import numpy as np
```

```
In [10]: arr = np.random.rand(4, 2)
```

```
In [11]: arr # 4 x 2 numpy array between 0 and 1
```

```
Out[11]: array([[ 0.31922229,  0.38354227],
 [ 0.22673472,  0.19127405],
 [ 0.55112153,  0.35564095],
 [ 0.11184216,  0.41846052]])
```

```
In [12]: pd.DataFrame(arr)
```

```
Out[12]:
```

	0	1
0	0.319222	0.383542
1	0.226735	0.191274
2	0.551122	0.355641
3	0.111842	0.418461

```
In [13]: pd.DataFrame(arr, columns = ['one', 'two'])
```

```
Out [13]:
```

	one	two
0	0.319222	0.383542
1	0.226735	0.191274
2	0.551122	0.355641
3	0.111842	0.418461

```
In [14]: pd.DataFrame({'student': np.arange(100, 110, 1), 'test': np.random.randint(
# create bigger dataset using randint and range
# arange like range => inclusive, exclusive, step
# randint => start, stop, total number of numbers you want
```

```
Out [14]:
```

	student	test
0	100	78
1	101	87
2	102	86
3	103	92
4	104	95
5	105	72
6	106	92
7	107	62
8	108	69
9	109	62

```
In [15]: pd.DataFrame({'student': np.arange(100, 110, 1), 'test': np.random.randint
```

```
Out [15]:
```

	test
student	
100	66
101	99
102	98
103	69
104	70
105	69
106	92
107	77
108	79
109	88

```
In [20]: s = pd.Series(['round', 'square', 'potato'], index=['c', 'b', 'a'], name='s'
s
```

```
Out [20]: c    round
b    square
a    potato
Name: shape, dtype: object
```

```
In [21]: pd.concat([df, s], axis = 1) # concat side by side => axis = 1 vs rows co
```

```
Out [21]:
```

	id	color	shape
a	100	red	potato

```
b 101 blue square
c 102  red  round
```

```
In [ ]:
```