



INTRO TO PYTHON FOR DATA SCIENCE

Numpy: Basic Statistics



Data analysis

- Get to know your data
- Little data → simply look at it
- Big data → ?



City-wide survey

```
In [1]: import numpy as np
```

```
In [2]: np_city = ... # Implementation left out
```

```
In [3]: np_city
```

```
Out[3]:
```

```
array([[ 1.64,  71.78],  
       [ 1.37,  63.35],  
       [ 1.6  ,  55.09],  
       ...,  
       [ 2.04,  74.85],  
       [ 2.04,  68.72],  
       [ 2.01,  73.57]])
```



Numpy

```
In [4]: np.mean(np_city[:,0])  
Out[4]: 1.7472
```

```
In [5]: np.median(np_city[:,0])  
Out[5]: 1.75
```

```
In [6]: np.corrcoef(np_city[:,0], np_city[:,1])  
Out[6]:  
array([[ 1.          , -0.01802],  
       [-0.01803,  1.          ]])
```

```
In [7]: np.std(np_city[:,0])  
Out[7]: 0.1992
```

- `sum()`, `sort()`, ...
- Enforce single data type: speed!



Generate data

distribution
mean

distribution
standard dev.

number of
samples

```
In [8]: height = np.round(np.random.normal(1.75, 0.20, 5000), 2)
```

```
In [9]: weight = np.round(np.random.normal(60.32, 15, 5000), 2)
```

```
In [10]: np_city = np.column_stack((height, weight))
```



INTRO TO PYTHON FOR DATA SCIENCE

Let's practice!