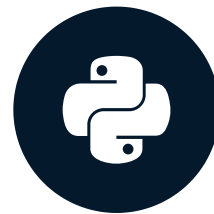


# Numpy

## INTRODUCTION TO PYTHON



**Hugo Bowne-Anderson**  
Data Scientist at DataCamp

# Lists Recap

- Powerful
- Collection of values
- Hold different types
- Change, add, remove
- Need for Data Science
  - Mathematical operations over collections
  - Speed

# Illustration

```
height = [1.73, 1.68, 1.71, 1.89, 1.79]  
height
```

```
[1.73, 1.68, 1.71, 1.89, 1.79]
```

```
weight = [65.4, 59.2, 63.6, 88.4, 68.7]  
weight
```

```
[65.4, 59.2, 63.6, 88.4, 68.7]
```

```
weight / height ** 2
```

```
TypeError: unsupported operand type(s) for **: 'list' and 'int'
```

# Solution: Numpy

- Numeric Python
- Alternative to Python List: Numpy Array
- Calculations over entire arrays
- Easy and Fast
- Installation
  - In the terminal: `pip3 install numpy`

# Numpy

```
import numpy as np
np_height = np.array(height)
np_height
```

```
array([ 1.73,  1.68,  1.71,  1.89,  1.79])
```

```
np_weight = np.array(weight)
np_weight
```

```
array([ 65.4,  59.2,  63.6,  88.4,  68.7])
```

```
bmi = np_weight / np_height ** 2
bmi
```

```
array([ 21.852,  20.975,  21.75 ,  24.747,  21.441])
```

# Comparison

```
height = [1.73, 1.68, 1.71, 1.89, 1.79]  
weight = [65.4, 59.2, 63.6, 88.4, 68.7]  
weight / height ** 2
```

```
TypeError: unsupported operand type(s) for **: 'list' and 'int'
```

```
np_height = np.array(height)  
np_weight = np.array(weight)  
np_weight / np_height ** 2
```

```
array([ 21.852,  20.975,  21.75 ,  24.747,  21.441])
```

# Numpy: remarks

```
np.array([1.0, "is", True])
```

```
array(['1.0', 'is', 'True'],  
      dtype='<U32')
```

- Numpy arrays: contain only one type

# Numpy: remarks

```
python_list = [1, 2, 3]  
numpy_array = np.array([1, 2, 3])
```

```
python_list + python_list
```

```
[1, 2, 3, 1, 2, 3]
```

```
numpy_array + numpy_array
```

```
array([2, 4, 6])
```

- Different types: different behavior!



# Numpy Subsetting

```
bmi
```

```
array([ 21.852,  20.975,  21.75 ,  24.747,  21.441])
```

```
bmi[1]
```

```
20.975
```

```
bmi > 23
```

```
array([False, False, False,  True, False], dtype=bool)
```

```
bmi[bmi > 23]
```

```
array([ 24.747])
```

# Let's practice!

INTRODUCTION TO PYTHON

# 2D Numpy Arrays

INTRODUCTION TO PYTHON



**Hugo Bowne-Anderson**  
Data Scientist at DataCamp

# Type of Numpy Arrays

```
import numpy as np
np_height = np.array([1.73, 1.68, 1.71, 1.89, 1.79])
np_weight = np.array([65.4, 59.2, 63.6, 88.4, 68.7])
```

```
type(np_height)
```

```
numpy.ndarray
```

```
type(np_weight)
```

```
numpy.ndarray
```

# 2D Numpy Arrays

```
np_2d = np.array([[1.73, 1.68, 1.71, 1.89, 1.79],  
                  [65.4, 59.2, 63.6, 88.4, 68.7]])  
  
np_2d
```

```
array([[1.73, 1.68, 1.71, 1.89, 1.79],  
       [65.4, 59.2, 63.6, 88.4, 68.7]])
```

```
np_2d.shape
```

```
(2, 5) # 2 rows, 5 columns
```

```
np.array([[1.73, 1.68, 1.71, 1.89, 1.79],  
          [65.4, 59.2, 63.6, 88.4, "68.7"]])
```

```
array([[ '1.73', '1.68', '1.71', '1.89', '1.79'],  
       [ '65.4', '59.2', '63.6', '88.4', '68.7']],  
      dtype='<U32')
```

# Subsetting

	0	1	2	3	4	
array([[	1.73,	1.68,	1.71,	1.89,	1.79],	0
[	65.4,	59.2,	63.6,	88.4,	68.7]])	1

```
np_2d[0]
```

```
array([ 1.73,  1.68,  1.71,  1.89,  1.79])
```

# Subsetting

	0	1	2	3	4	
array([[	1.73,	1.68,	1.71,	1.89,	1.79],	0
[	65.4,	59.2,	63.6,	88.4,	68.7]])	1

```
np_2d[0][2]
```

```
1.71
```

```
np_2d[0,2]
```

```
1.71
```

# Subsetting

	0	1	2	3	4	
array([[	1.73,	1.68,	1.71,	1.89,	1.79],	0
[	65.4,	59.2,	63.6,	88.4,	68.7]])	1

```
np_2d[:,1:3]
```

```
array([[ 1.68,  1.71],  
       [ 59.2 , 63.6 ]])
```

```
np_2d[1,:]
```

```
array([ 65.4,  59.2,  63.6,  88.4,  68.7])
```



# Let's practice!

INTRODUCTION TO PYTHON

# Numpy: Basic Statistics

INTRODUCTION TO PYTHON



**Hugo Bowne-Anderson**  
Data Scientist at DataCamp

# Data analysis

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

# City-wide survey

```
import numpy as np
np_city = ... # Implementation left out
np_city
```

```
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

```
np.mean(np_city[:,0])
```

```
1.7472
```

```
np.median(np_city[:,0])
```

```
1.75
```

# Numpy

```
np.corrcoef(np_city[:,0], np_city[:,1])
```

```
array([[ 1.        , -0.01802],  
       [-0.01803,  1.        ]])
```

```
np.std(np_city[:,0])
```

```
0.1992
```

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

# Generate data

- Arguments for `np.random.normal()`
  - distribution mean
  - distribution standard deviation
  - number of samples

```
height = np.round(np.random.normal(1.75, 0.20, 5000), 2)

weight = np.round(np.random.normal(60.32, 15, 5000), 2)

np_city = np.column_stack((height, weight))
```

# Let's practice!

INTRODUCTION TO PYTHON