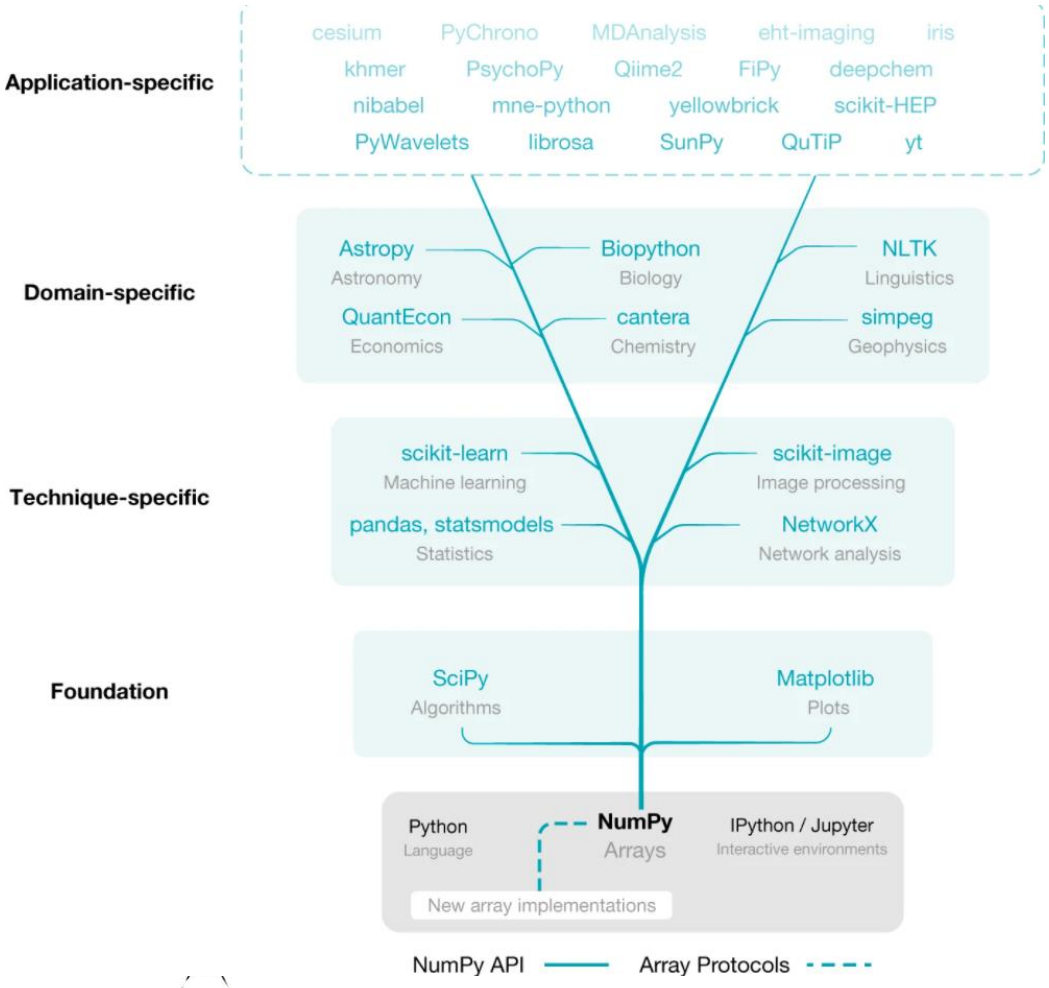
INTRODUCTION TO DATA SCIENCE

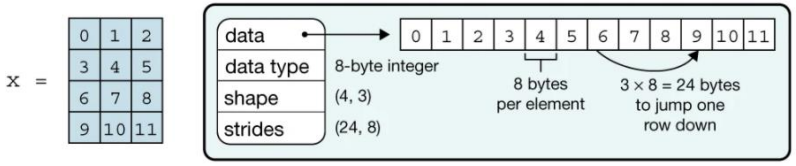
TA class III – Numpy

TA : Lee Chi-Hsuan

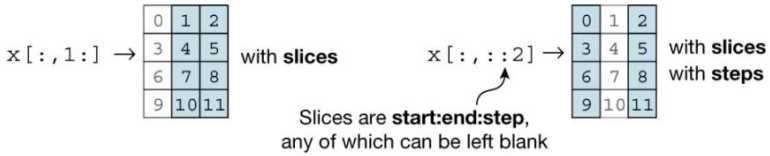
numpy


Application-specific
cesium
PyChrono
MDAnalysis
eht-imaging
iris
khmer
PsychoPy
Qiime2
FiPy
deepchem
nibabel
mne-python
yellowbrick
scikit-HEP
PyWavelets
librosa
SunPy
QuTiP
yt
Domain-specific
Astropy
Astronomy
Biopython
Biology
NLTK
Linguistics
QuantEcon
Economics
cantera
Chemistry
simpeg
Geophysics
Technique-specific
scikit-learn
Machine learning
scikit-image
Image processing
pandas, statsmodels
Statistics
NetworkX
Network analysis
Foundation
SciPy
Algorithms
Matplotlib
Plots
Python
Language
NumPy
Arrays
IPython / Jupyter
Interactive environments
New array implementations
NumPy API
Array Protocols

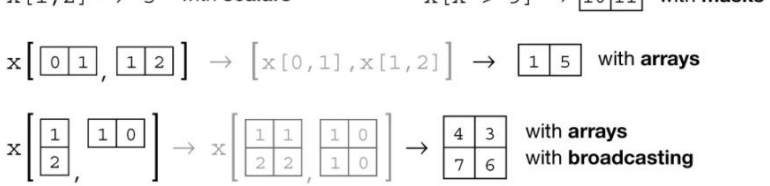
a Data structure



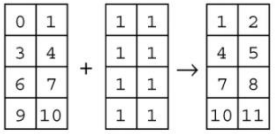
b Indexing (view)



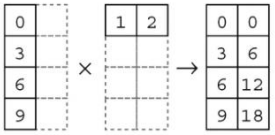
c Indexing (copy)



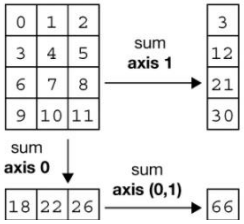
d Vectorization



e Broadcasting



f Reduction

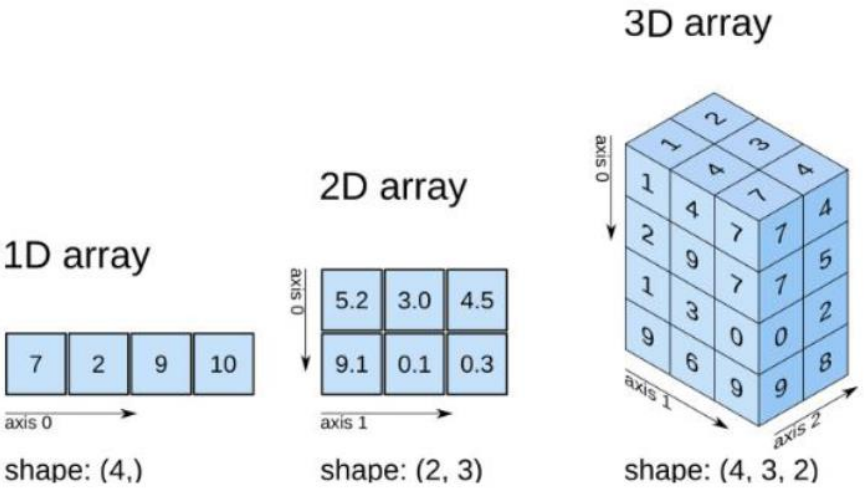


g Example

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

In [2]: x = np.arange(12)

In [3]: x = x.reshape(4, 3)

In [4]: x
Out[4]:
array([[ 0,  1,  2],
       [ 3,  4,  5],
       [ 6,  7,  8],
       [ 9, 10, 11]])

In [5]: np.mean(x, axis=0)
Out[5]: array([4.5, 5.5, 6.5])

In [6]: x = x - np.mean(x, axis=0)

In [7]: x
Out[7]:
array([[-4.5, -4.5, -4.5],
       [-1.5, -1.5, -1.5],
       [ 1.5,  1.5,  1.5],
       [ 4.5,  4.5,  4.5]])
```

key features :

- ndim
- shape
- dtype



- Some important concepts…

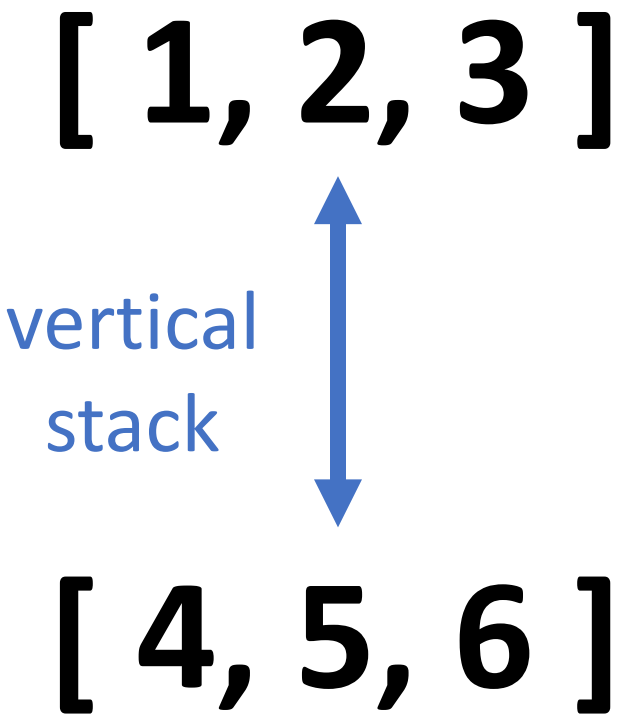
vstack vs. hstack

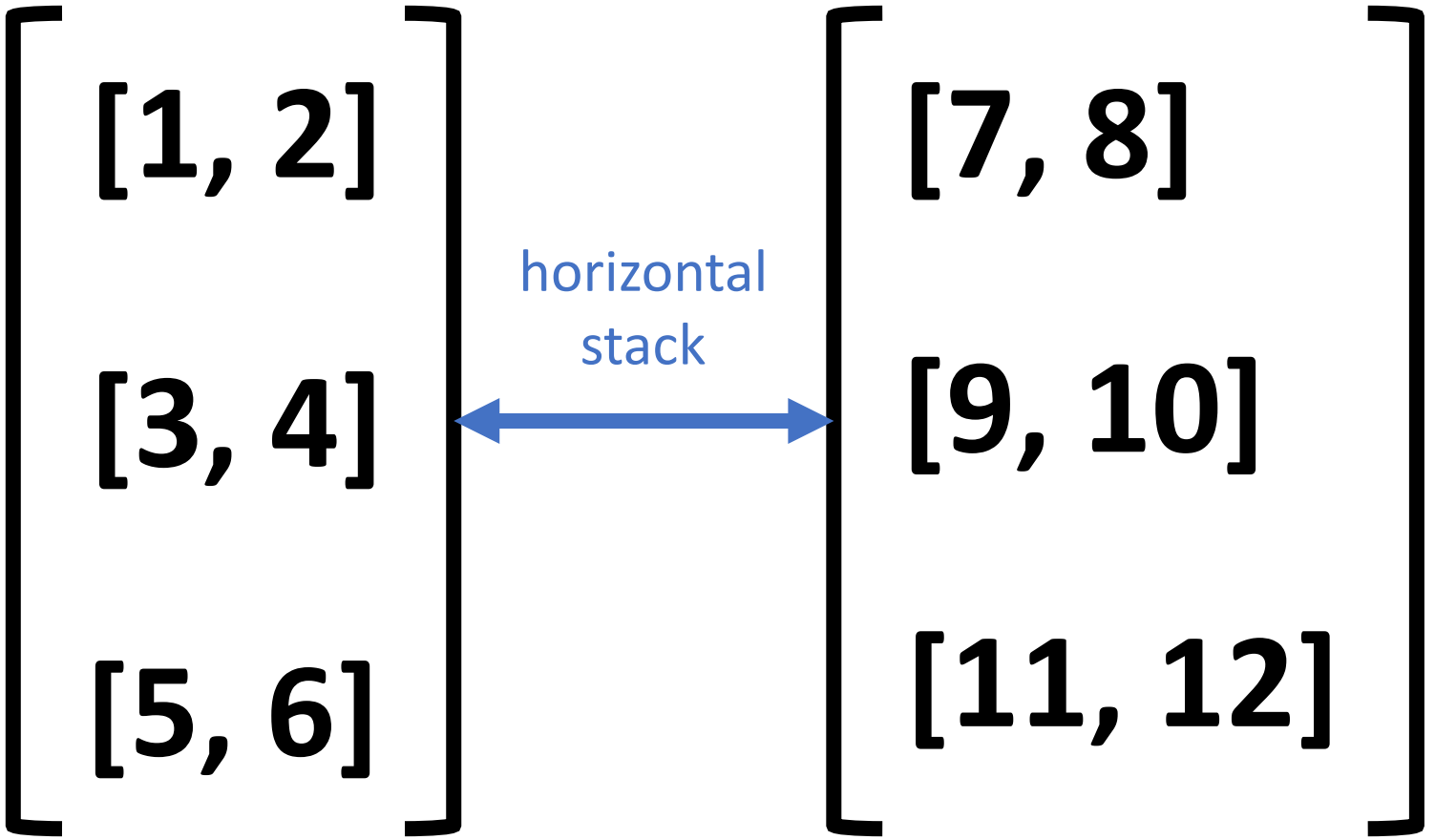
```
arr1 = np.array([1,2,3])
arr2 = np.array([4,5,6])
np.vstack((arr1, arr2))
```

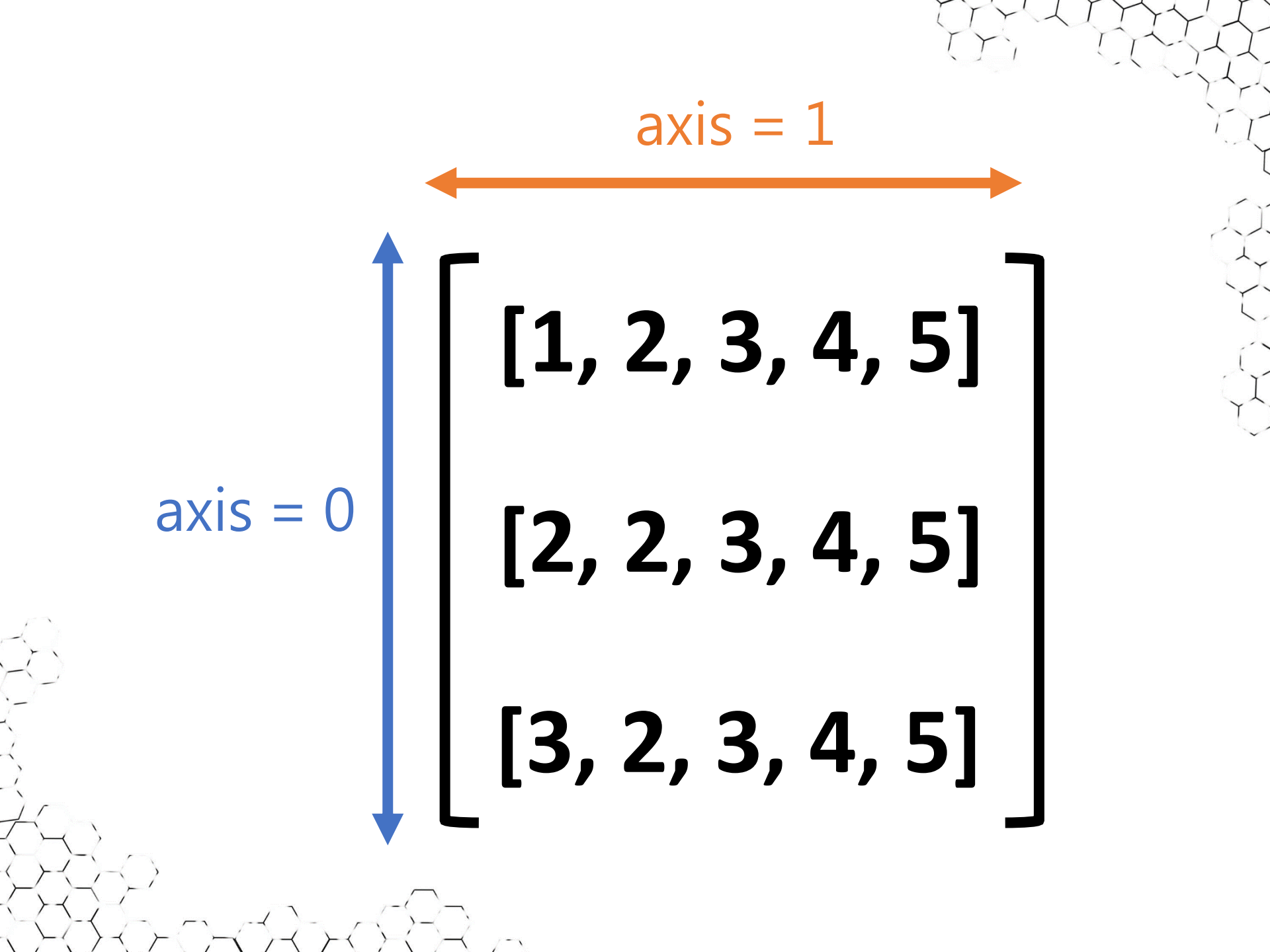
```
array([[1, 2, 3],
       [4, 5, 6]])
```

```
arr3 = np.array([[1,2], [3,4], [5,6]])
arr4 = np.array([[7,8], [9,10], [11,12]])
np.hstack((arr3, arr4))
```

```
array([[ 1,  2,  7,  8],
       [ 3,  4,  9, 10],
       [ 5,  6, 11, 12]])
```


[1, 2, 3]
vertical
stack
[4, 5, 6]


[1, 2]
[3, 4]
[5, 6]
horizontal
stack
[7, 8]
[9, 10]
[11, 12]


axis = 1
axis = 0
[1, 2, 3, 4, 5]
[2, 2, 3, 4, 5]
[3, 2, 3, 4, 5]

```
# arr = [[1,2],
#        [3,4],
#        [5,6]]

arr = np.array([[1,2], [3,4], [5,6]])
print('sum(arr, axis = 0) = ' + str(np.sum(arr, axis = 0)))
print('sum(arr, axis = 1) = ' + str(np.sum(arr, axis = 1)))
```

```
sum(arr, axis = 0) = [ 9 12]
sum(arr, axis = 1) = [ 3  7 11]
```

Let's have a quick walkthrough

Quick numpy walkthrough

```
import numpy as np
```

```
n = np.zeros(10)
```