

SIC Batch 5

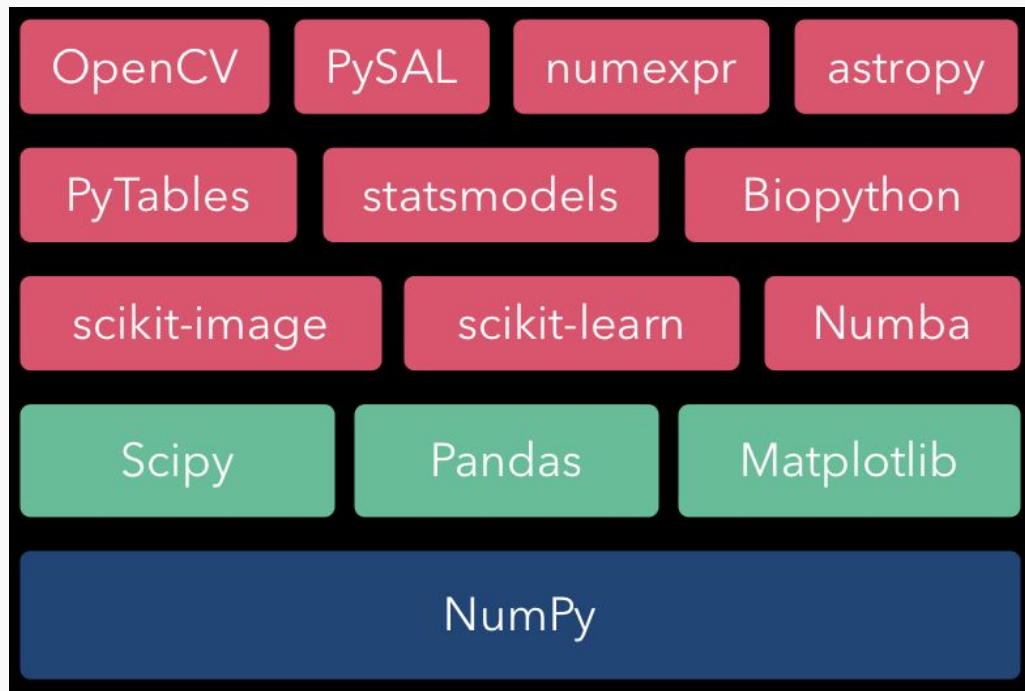
Week 2 - Python for AI & Data Science

What is Numpy

NumPy stands for 'Numerical Python'. It is a package for data analysis and scientific computing with Python. NumPy uses a multidimensional array object, and has functions and tools for working with these arrays. The powerful n-dimensional array in NumPy speeds-up data processing.



**NumPy is
the
foundation
of the
python
scientific
stack**



What is Numpy

NumPy is suited to many applications

- Image processing
- Signal processing
- Linear algebra
- A plethora of others

What is Numpy

NumPy provides **vectorization** of mathematical operations on arrays and matrices which significantly improves the performance.

Arrays are important because they enable you to express **batch operations** on data **without writing any for loops**. This practice of replacing explicit loops with array expressions is commonly referred to as **vectorization**.

```
In [51]: arr = np.array([[1.,  
2., 3.], [4., 5., 6.]])
```

```
In [52]: arr  
Out[52]:  
array([[1., 2., 3.],  
       [4., 5., 6.]])
```

```
In [53]: arr * arr  
Out[53]:  
array([[ 1.,  4.,  9.],  
       [16., 25., 36.]])
```

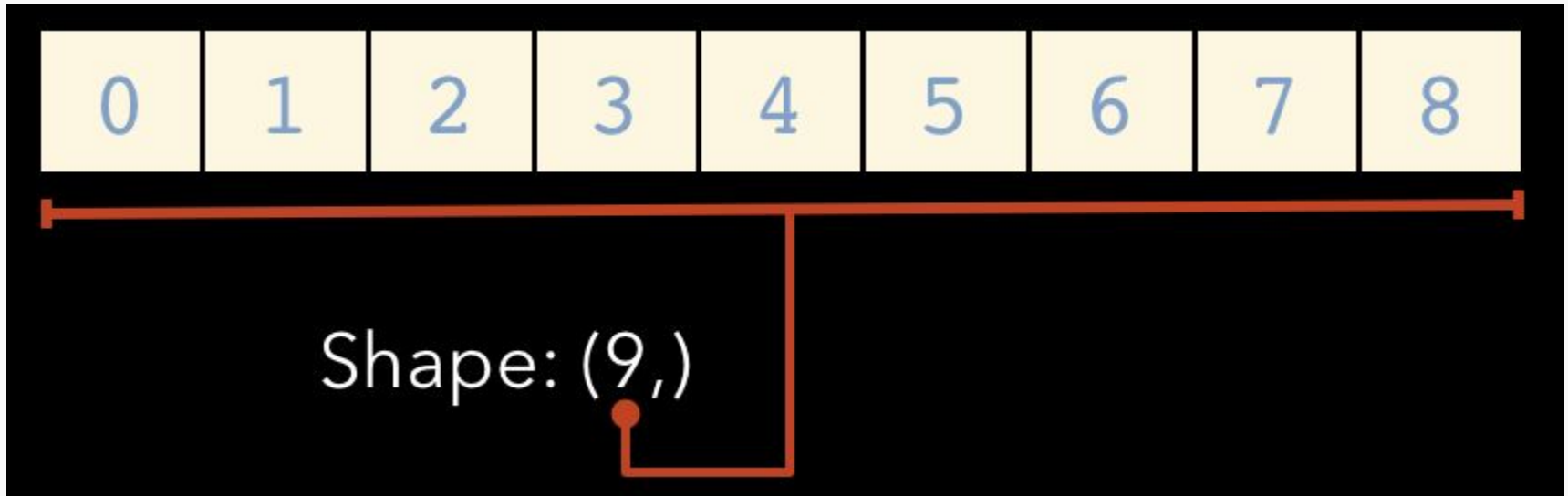
```
In [54]: arr - arr  
Out[54]:  
array([[0., 0., 0.],  
       [0., 0., 0.]])
```

Array

An array is a **data type** used to store multiple values using a single identifier (variable name). An array contains an **ordered** collection of data elements where each element is of the **same type** and can be referenced by its **index** (position).

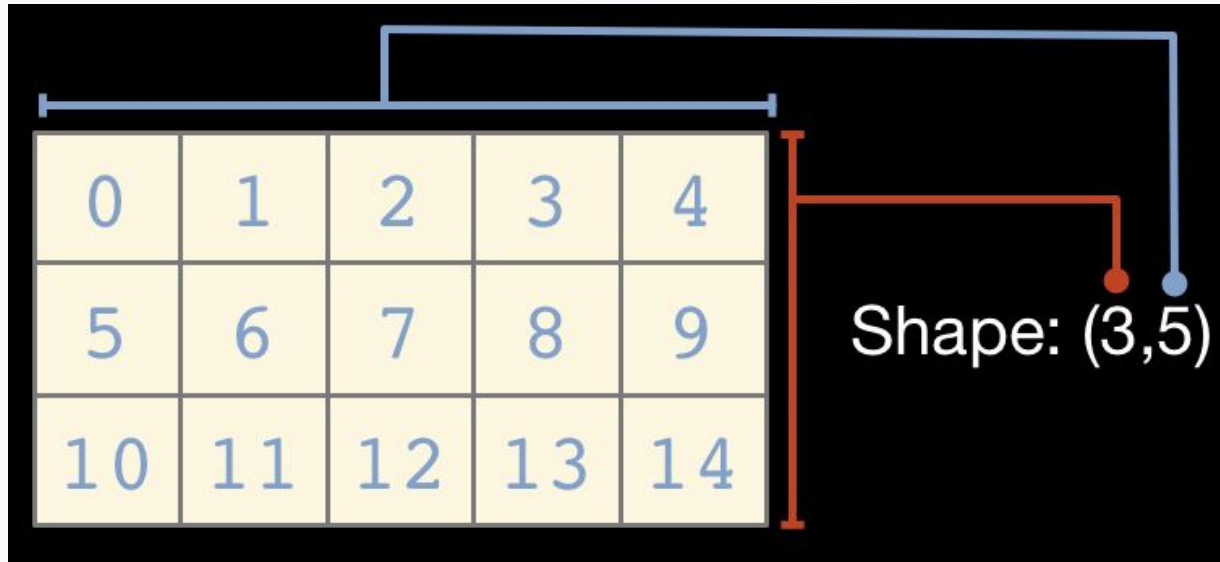
Array Dimension

One dimensional arrays have a 1-tuple for their shape



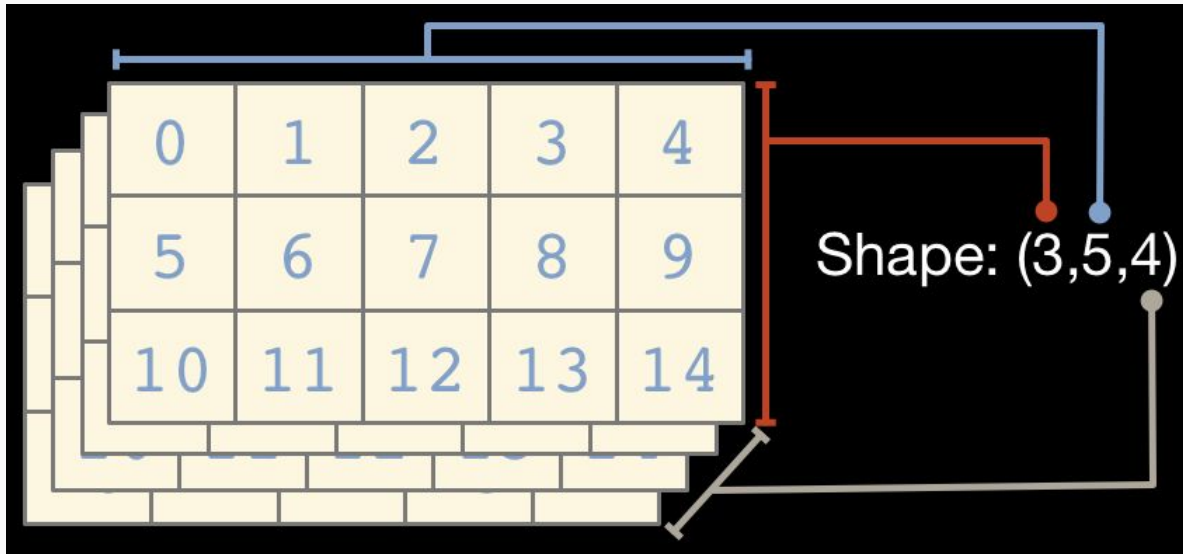
Array Dimension

...Two dimensional arrays have a 2-tuple



Array Dimension

...And so on



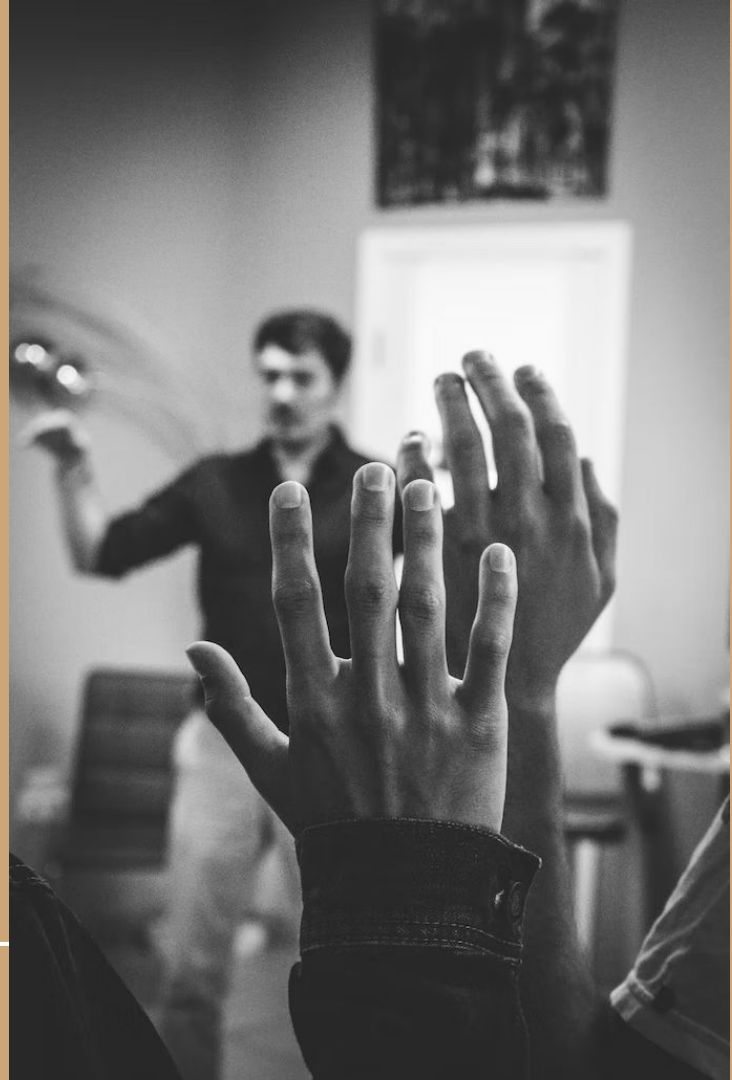
Difference Between List and Array

List	Array
List can have elements of different data types for example, [1, 3.4, 'hello', 'a@']	All elements of an array are of same data type for example, an array of floats may be: [1.2, 5.4, 2.7]
Elements of a list are not stored contiguously in memory.	Array elements are stored in contiguous memory locations . This makes operations on arrays faster than lists.
Lists do not support element wise operations , for example, addition, multiplication, etc. because elements may not be of same type.	Arrays support element wise operations . For example, if A1 is an array, it is possible to say A1/3 to divide each element of the array by 3.

Difference Between List and Array

List	Array
<p>Lists can contain objects of different data type that Python must store the type information for every element along with its element value. Thus lists take more space in memory and are less efficient.</p>	<p>NumPy array takes up less space in memory as compared to a list because arrays do not require to store datatype of each element separately.</p>
<p>List is a part of core Python.</p>	<p>Array (ndarray) is a part of NumPy library.</p>

Challenges



Series vs Dataframe

- A Series is essentially a **column**,
- A DataFrame is a multi-dimensional **table** made up of a collection of Series.

Series

	apples
0	3
1	2
2	0
3	1

+

Series

	oranges
0	0
1	3
2	7
3	2

=

DataFrame

	apples	oranges
0	3	0
1	2	3
2	0	7
3	1	2

What is Pandas

Pandas is a Python library used for working with data sets. It has functions for analyzing, cleaning, exploring, and manipulating data.

- Adds data structures and tools designed to work with table-like data (similar to Series and Data Frames in R)
- Provides tools for data manipulation: reshaping, merging, sorting, slicing, aggregation etc.
- Allows handling missing data



Series

A Series is a one-dimensional array-like object containing a sequence of values (of similar types to NumPy types) and an associated array of data labels, called its index

```
In [11]: obj = pd.Series([4, 7, -5, 3])
```

```
In [12]: obj
```

```
Out[12]:
```

```
0    4
```

```
1    7
```

```
2   -5
```

```
3    3
```

```
dtype: int64
```

0	4
1	7
2	-5
3	3

DataFrame

A DataFrame represents a rectangular table of data and contains an ordered collection of columns, each of which can be a different value type (numeric, string, boolean, etc.). The DataFrame has both a row and column index; it can be thought of as a dict of Series all sharing the same index.

```
data = {'state': ['Ohio', 'Ohio', 'Ohio', 'Nevada', 'Nevada',  
                'Nevada'],
```

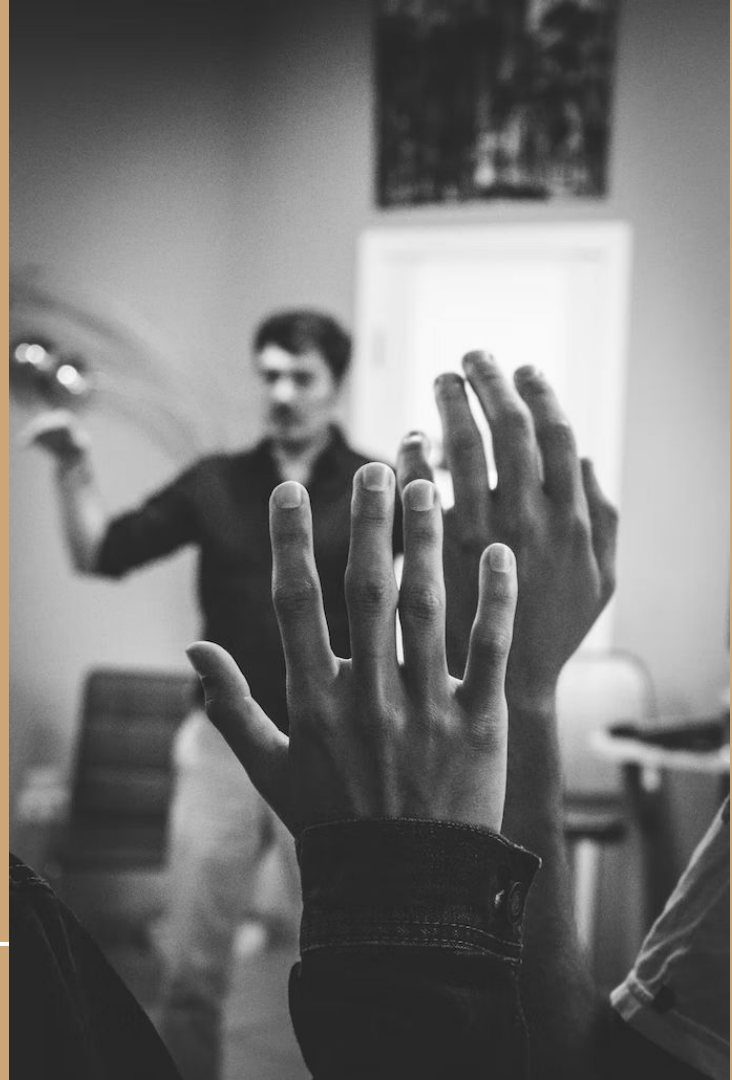
```
        'year': [2000, 2001, 2002, 2001, 2002, 2003],
```

```
        'pop': [1.5, 1.7, 3.6, 2.4, 2.9, 3.2]}
```

```
frame = pd.DataFrame(data)
```

	pop	state	year
0	1.5	Ohio	2000
1	1.7	Ohio	2001
2	3.6	Ohio	2002
3	2.4	Nevada	2001
4	2.9	Nevada	2002

Challenges



Why we need Visualization

Making informative visualizations (sometimes called plots) is one of the most important tasks in data analysis. It may be a part of the exploratory process—for example, to help identify outliers or needed data transformations, or as a way of generating ideas for models.



Matplotlib

- Matplotlib is the most popular Python library designed for creating (mostly two-dimensional) publication-quality plots.
- The project was started by John Hunter in 2002 to enable a MATLAB-like plotting interface in Python.
- This library is built on the top of NumPy arrays and consist of several plots like line chart, bar chart, histogram, etc. it also can export visualizations to all of the common vector and raster graphics formats (PDF, SVG, JPG, PNG, BMP, GIF, etc.).



Challenges

