



CS196

Data Science

Numpy and Pandas



Week 5

One Line:

```
print(list(map(lambda filename: [i['lead_paragraph'] for i in json.load(open(filename))['response']['docs']],
glob.glob("*.json"))))
```

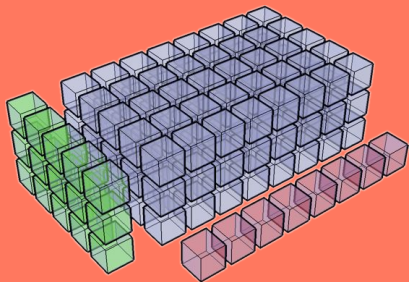
Four Lines:

```
for file in glob.glob('<path>'):
    with open(file) as fileText: jsonText = json.loads(fileText.read())
    for i in range(0, len(jsonText['response']['docs']):
        print(jsonText['response']['docs'][i]['lead_paragraph'])
```

Numpy

Numerical Python...

There is also SciPy,
which is Scientific Python for scientific
computing.

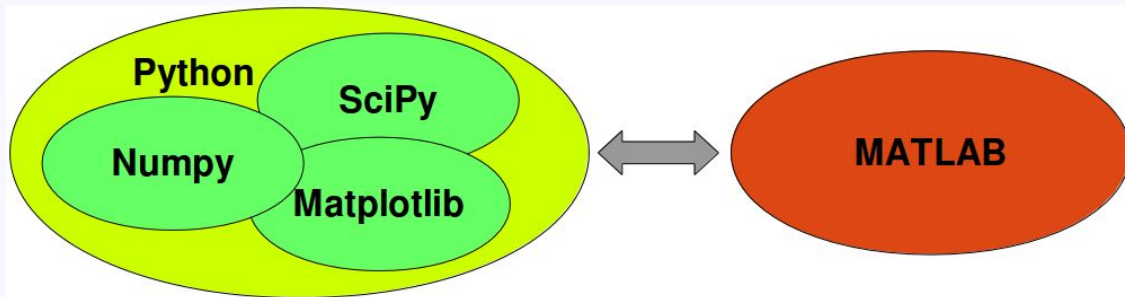


Advantages of Core Python:

- High-level number objects: integers, floating point
- Containers: List with cheap insertion and append methods, dictionaries with fast lookup

Advantages of using Numpy with Python:

- Array Oriented Computing
- Efficiently implemented multi-dimensional arrays
- Designed for scientific computation



Universal Functions (numpy)

- “Ufuncs” operate on every element in an array
- Allows you to abstract away having to loop through arrays
- Many, many functions supported
 - Arithmetic, trigonometry, bitwise, etc.

"How is NumPy so fast?"

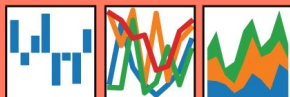
- Arrays are typed with a single data type
 - Allows easier lookups and data packing
- NumPy Arrays are actually Arrays (not Lists)
 - Much close to C arrays

Pandas

DATA FRAME!!!

pandas

$$y_i t = \beta' x_{it} + \mu_i + \epsilon_{it}$$

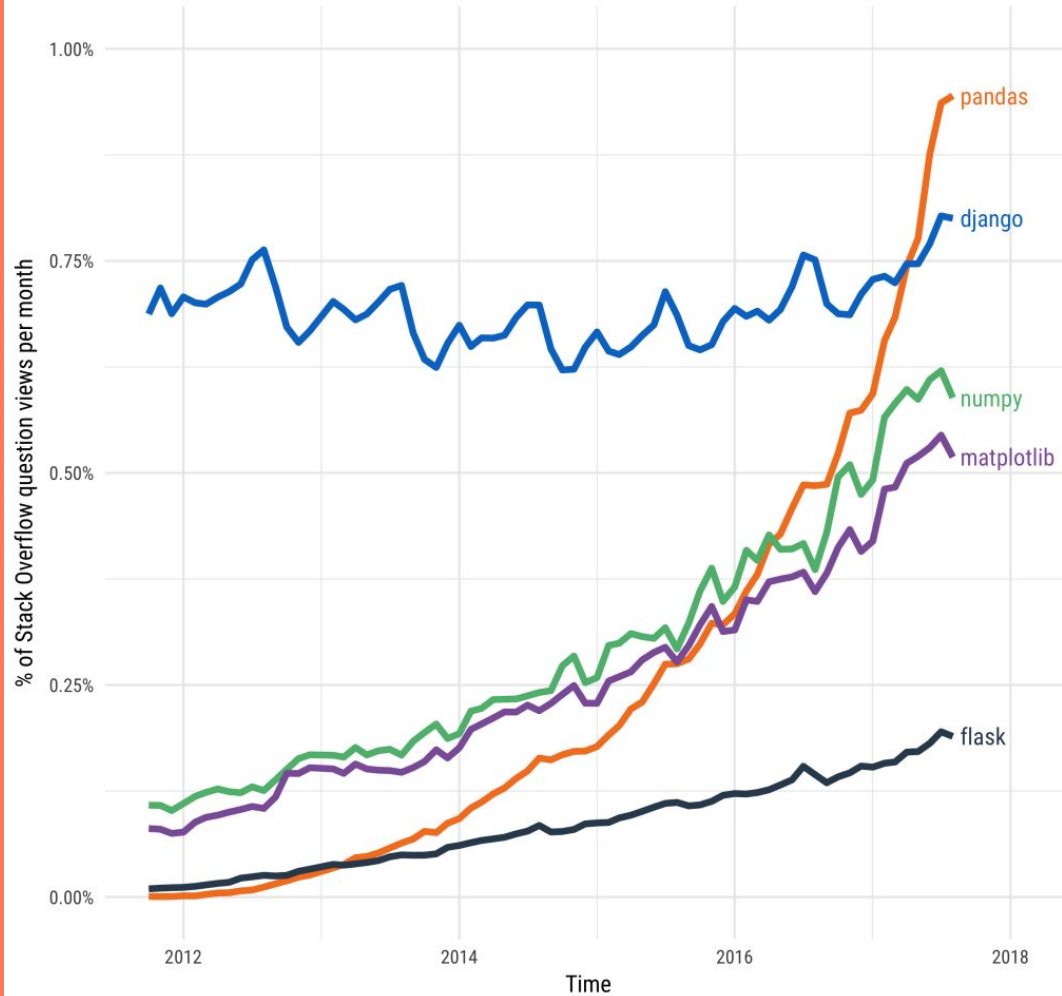


When to use Pandas :

- Indexing
- Renaming
- Handling Missing Values
- map(), apply(), applymap()
- groupby()
- New Columns = f(Existing Columns)
- Basic Stats
- Merge, join
- Simple Plots
- Etc.

Stack Overflow Traffic to Questions About Selected Python Packages

Based on visits to Stack Overflow questions from World Bank high-income countries



conda install numpy pandas

Jupyter Time

Challenge

Weekly challenges build on top of each other and will get progressively harder.

Stay on top of them.

- Weekly Challenges will slowly develop into Data Science technical interview questions
- If you are struggling, come talk to us after each Hackerspace, and/or you can always ping us on slack for help.
- Google is your best friend.
- Have fun and enjoy the ride :)



