CS196

Data Science

CS196

Numpy and Pandas

(ft. Not Ben)

Week 4 of 196 (week 5 of school)

Ben has a midterm :(

Channel: #data_hackerspace

Current Leaderboard

1st Place: Osmar Coronel

2nd Place: Drake Eidukas
Ish Shah
Matt Cannalte
Tyson Traugar

3rd Place: David Brewster
Justin Chan
Dean Lin

You can still get points by sending us the answers after before the next hackerspace

Week 3 Challenge Answers

One liner by Osmar:

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

Two liner by Dean:

```
for filename in glob.glob('*.json'):
    print json.loads(open(filename).read())['response']['docs'][0]['lead_paragraph']
```

Four liner by Aaron:

```
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']
```

3rd Place

Amazon Echo Dot

X 3

2nd Place

Fitbit Time 2

Or

Pebble 2

1st Place

Mini Jambox

4th Place

The best prize everrr

Selfie Opportunity with Tyler and Ben

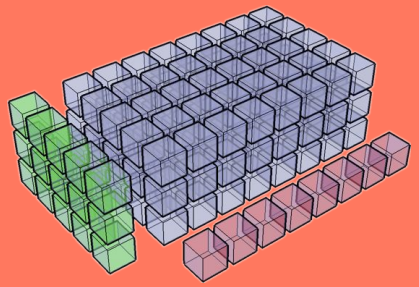
AND/OR

Opportunity to take a GenEd course with Tyler next semester

Numpy

Numerical Python...

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

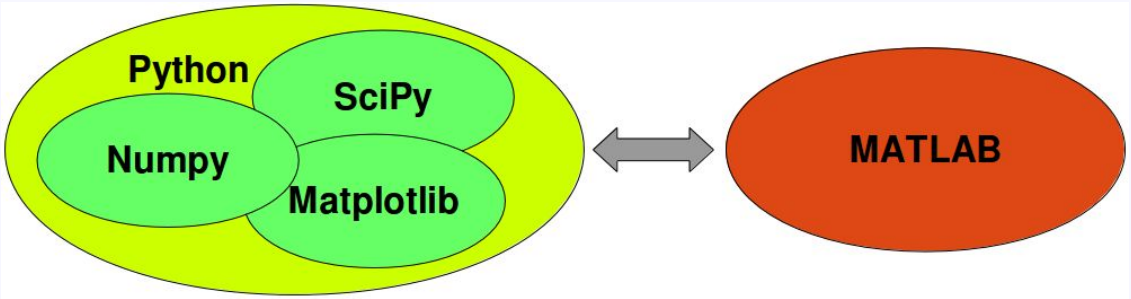


Advantages of Core Python:

- 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



Pandas

DATA FRAME!!!

I personally don't use Pandas but-
When to use Pandas :

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

Jupyter Time

http://tinyurl.com/besthackerspace

#praiseBen

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.
- Late Submission is a thing.
- You need to complete more than 70% of all challenges to get an A in this hackerspace.
- Answers to each week's challenge will be either sent out through emails or announced at next Hackerspace (Depends on how lazy I am that week).
- Google is your best friend.
- Have fun and enjoy the ride :)