Python for EOAS

Examples of using modules

Learning Goals

1. Read CSV data from files into NumPy data structures, using pandas
2. Use list boolean slices to select data
3. Use requests to get data from the web

Get the Data

- Use your browser to go to `http://climate.weather.gc.ca/` and work your way through to the "Hourly Data Report" for yesterday at the *Vancouver Intl A* station.
- Download the August 2013 hourly data as a CSV file
- Use your shell skills to confirm that:
  - You really got a CSV file
  - It's for the *Vancouver Intl A* station
  - It contains hourly data for the whole month of August 2013
- Move or copy the CSV file into the *data-explore/* directory in your repo and commit it.

1. Read CSV data from files into NumPy data structures, using pandas

- import numpy as np
- import pandas as pd
- !head eng-hourly-08012013-08312013.csv
- !head -20 eng-hourly-08012013-08312013.csv
- !tail eng-hourly-08012013-08312013.csv
- data = pd.read_csv('eng-hourly-08012013-08312013.csv', skiprows=16)
- print data[0:4]
- print data.tail(1)
- !tail -1 eng-hourly-08012013-08312013.csv
- data = pd.read_csv('eng-hourly-08012013-08312013.csv', skiprows=16, encoding="ISO-8859-1")
- print data.columns

2. Use list boolean slices to select data

- temps = data[u'Temp (C)']
- print 'max:', temps.max(), 'on', data['Date/Time'][temps.argmax()]
- print 'min:', temps.min(), 'on', data['Date/Time'][temps.argmin()]
- print 'mean:', temps.mean()
- print 'std dev:', temps.std()
- ```
  for day in range(1, 32):
      mask = data['Day']==day
      max_temp = temps[mask].max()
      date = data[mask]['Date/Time'][temps[mask].argmax()][:11]
      hour = data[mask]['Time'][temps[mask].argmax()]
      print 'max temperature on',date, 'was', max_temp, 'at', hr
  ```
  ```

Exercise

Plot the daily maximum temperature.

3. Use requests to get data from the web

- import requests
- url = 'http://climate.weather.gc.ca/climateData/bulkdata_e.html'
- params = {

```
    'timeframe': 1,
    'stationID': 51442,
    'Year': 2013,
    'Month': 7,
    'Day': 1,
    'format': 'csv',
}
```

- response = requests.get(url, params=params)
- response.headers
- from StringIO import StringIO
- fakefile = StringIO(response.content)
- datajul = pd.read_csv(fakefile, skiprows=16, encoding="ISO-8859-1")
- print datajul.head(2)

Exercise

Plot the daily maximum temperature for both August and July