Lecture 3: Importing data with Pandas

อ.ปรัชญ์ ปิยะวงศ์วิศาล

Today's Topics

Lab Exercise

- Importing data with Pandas

Download SMS Spam Dataset

- Download SMS spam data from
  - https://archive.ics.uci.edu/ml/datasets/SMS+Spam+Collection
  - Click "Data Folder"
- Unzip into work directory

Importing data the traditional way

Using open():

    with open('./SMSSpamCollection') as f:

        messages = f.readlines()

A more compact alternative:

    lines = [line.rstrip('\n') for line in open('filename')]

Importing data with Pandas

```python
messages = pd.read_csv('./SMSSpamCollection',
                       sep = '\t',
                       names = ["label","message"])
```

Playing with the data

```python
print(messages)

print(message.head())

messages.groupby('label').describe()


messages['length'] = messages['message'].map(lambda text: len(text))

print(messages.head())

messages.length.plot(bins=20, kind='hist')

messages.length.describe()
```