PYTHON FOR R USERS

Load up and look at some data

Daniel Chen
Instructor

Overview

- 2013 NYC Flights

- Load and Explore

- Manipulate and plot

Importing data with List Comprehensions

- List comprehension to load data

- Saves repetitive typing

- Condenced way of appending values to a list

List Comprehensions

```
In [1]: import glob

In [2]: import pandas as pd

In [3]: csv_files = glob.glob('*.csv')

In [4]: csv_files
Out[4]: ['data3.csv', 'data2.csv', 'data1.csv']

In [5]: all_dfs = [pd.read_csv(x) for x in csv_files]

In [6]: all_dfs[0]
Out[6]:
    A   B   C   D
0  a0  b0  c0  d0
1  a1  b1  c1  d1
2  a2  b2  c2  d2
3  a3  b3  c3  d3
```

PYTHON FOR R USERS

Let's practice!

PYTHON FOR R USERS

Manipulating data

Daniel Chen
Instructor

Groupby

```
In [68]: df = pd.DataFrame({
    ...:        'name':['John Smith', 'Jane Doe', 'Mary Johnson'],
    ...:        'treatment_a': [np.NaN, 16, 3],
    ...:        'treatment_b': [2, 11, 1]
    ...: })
    ...: df_melt = pd.melt(df, id_vars='name')
    ...:

In [69]: df_melt.groupby('name')['value'].mean()
Out[69]:
name
Jane Doe          13.5
John Smith         2.0
Mary Johnson       2.0
Name: value, dtype: float64
```

Groupby agg aggregate

```
In [70]: df_melt.groupby('name')['value'].agg(['mean', 'max'])
Out[70]:
                mean    max

name
Jane Doe        13.5   16.0
John Smith       2.0    2.0
Mary Johnson     2.0    3.0
```

Dummy Variables

- Categorical variables need to be encoded as dummy variables

- One-hot encoding

```
In [21]: df = pd.DataFrame({
    ...:         'status':['sick', 'healthy', 'sick'],
    ...:         'treatment_a': [np.NaN, 16, 3],
    ...:         'treatment_b': [2, 11, 1]
    ...:  })
    ...: df
    ...:
Out[21]:
    status  treatment_a  treatment_b
0     sick          NaN            2
1  healthy         16.0           11
2     sick          3.0            1
```

Get Dummies

```
In [5]: pd.get_dummies(df)
Out[5]:
   treatment_a   treatment_b   status_healthy   status_sick
0        NaN             2                 0             1
1       16.0            11                 1             0
2        3.0             1                 0             1
```

PYTHON FOR R USERS

Let's practice!

PYTHON FOR R USERS

Wrap-up

Daniel Chen
Instructor

Review of topics

- How R translates into Python

- Basic types, functions, methods

- Numpy and Pandas

- Data Manipulation and cleaning techniques

- Visualization

R vs Python

- One language isn't "better" than the other

- Broaden your toolkit

PYTHON FOR R USERS

Let's practice!