

Pandas Data Frames

Presenter: Steve Baskauf
steve.baskauf@vanderbilt.edu



Jean & Alexander Heard
LIBRARIES

CodeGraf landing page

- vanderbi.it/codegraf

What is Pandas ?

Pandas library for Python

- Open source library for data analysis
- Name from "panel data" (econometrics term)
- Built on NumPy
- Creates a DataFrame object
- Easy import from CSV and Excel

Import statement

- Conventional format:

```
import pandas as pd
```

- Use this to make your code compatible with everyone else's code

Series



Jean & Alexander Heard
LIBRARIES

Series

- A specific one-dimensional data structure with **named elements**.
- Pandas series are built from NumPy arrays + an index (numbers **and labels**).
- Series items can be referenced by either their integer index or string label.
- Series are built by instantiating a **pd.Series()** object
- Use **index = [list]** argument to assign a list as the labels.
- Series can be used in vectorized operations just like NumPy arrays.

Data frames

Data frames

- A specific two-dimensional data structure designed to be like a **spreadsheet**.
- Pandas data frames are built from a **set of series** (one series for each column).
- Column series share a **common index** (integer and label).
- Data frames are built by instantiating a **pd.DataFrame()** object.
- Use **index = [list]** argument to assign a list as the labels.

Data frames

data frame named `organism_info`

The diagram illustrates a pandas DataFrame named `organism_info`. It features a table with three columns: `index`, `group (0)`, and `number legs (1)`. The rows represent different organisms: 'lizard' (0), 'spider' (1), 'worm' (2), and 'bee' (3). Annotations include: a blue arrow pointing to the column label 'group (0)'; a black arrow pointing to the index label 'index'; a red arrow pointing to the 'group' column, labeled 'series'; and two black arrows pointing to the 'worm' row, labeled with `organism_info.loc['worm']` and `organism_info.iloc[2]`. Additionally, two black arrows point to the 'bee' row, labeled with `organism_info.iat[3,1]` and `organism_info.at['bee', 'number legs']`.

index	group (0)	number legs (1)
'lizard' (0)	'reptile'	4
'spider' (1)	'arachnid'	8
'worm' (2)	'annelid'	0
'bee' (3)	'insect'	6

range index

column label

index label

series

`organism_info.loc['worm']`
`organism_info.iloc[2]`

`organism_info['group']`
`organism_info.group`

`organism_info.iat[3,1]`
`organism_info.at['bee', 'number legs']`

- **Data frames** are essentially tables
- The **column values** are **series**.

Data frame attributes and methods

- `.head()` to display first five lines
- `.loc[]` to refer to a row by label
- `.iloc[]` to refer to a row by integer number
- `.index` to refer to the line indices

Loading data frames from files

File read and write functions

`pd.read_csv()` read from a CSV file into a data frame.

`pd.to_csv()` write from a data frame to a CSV file.

`pd.read_excel()` read from an Excel file into a data frame.

`pd.to_excel()` write from a data frame to an Excel file.

The read functions can be performed using a web URL

Remote Support for Teaching and Research Needs

Jean & Alexander Heard
LIBRARIES



Access to digital collections 24/7



Skype consultations with your
subject librarian



Ask a Librarian: an easy way to
submit a question via email



Live chat available from the
Library home page

NEED HELP? ASK A LIBRARIAN!

<https://www.library.vanderbilt.edu/ask-librarian.php>

Jean & Alexander Heard
LIBRARIES