Pandas Data Frames

Presenter: Steve Baskauf
steve.baskauf@vanderbilt.edu

DISC DIGITAL SCHOLARSHIP AND COMMUNICATIONS

Jean & Alexander Heard LIBRARIES

CodeGraf landing page

- vanderbi.lt/codegraf

What is Pandas ?

Jean & Alexander Heard
LIBRARIES

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

DISC DIGITAL SCHOLARSHIP AND COMMUNICATIONS

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

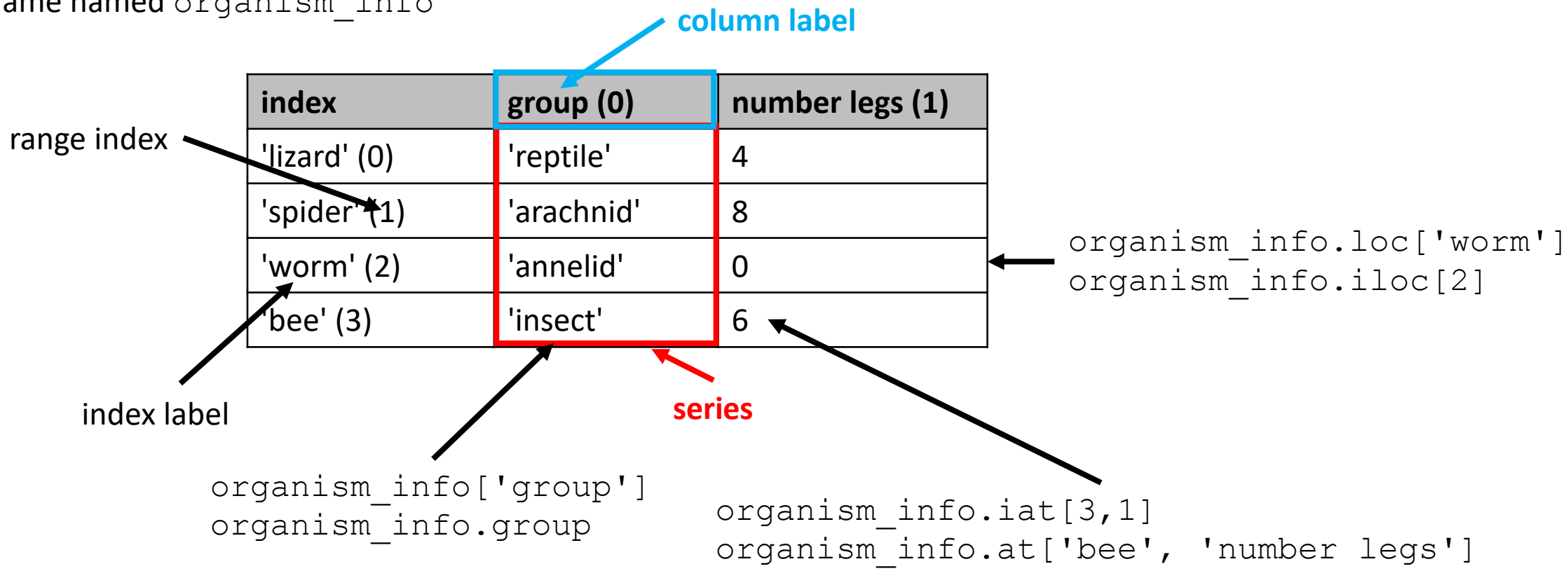
Jean & Alexander Heard
LIBRARIES

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`



- **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

Jean & Alexander Heard
LIBRARIES

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

- 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

Jean & Alexander Heard
LIBRARIES

NEED HELP? ASK A LIBRARIAN!

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

Jean & Alexander Heard
LIBRARIES