

COMP 333 — Week 4 pandas Dataframes

pandas Dataframes

Here we look at the features of Python **pandas**, especially the main datatype **dataframe**.

pandas was designed for Data Analytics.

pandas provides *labeled arrays*,
that is, arrays indexed by what I will call “headers”
to handle large scale data in a convenient way for
Descriptive Data Analysis, Data Wrangling, and Exploratory Data Analysis.

series is the 1-D labeled array datatype in **pandas**

dataframe is the 2-D labeled array datatype in **pandas**

xarray is the N-D labeled array datatype in **pandas**

pandas provides operations to summarize, filter, transform and reshape dataframes.

pandas integrates with the Python ecosystem **numpy**, **scipy**, **matplotlib** etc.

Chapter 5 of the **pandas** book provides an introduction to **pandas**, similar to this lecture.

We will cover

- ▶ Introduction to Python array datatype
- ▶ Introduction to **numpy ndarray** datatype
- ▶ Introduction to **series**, **dataframe**, **xarray** datatype
- ▶ **pandas** constructors for **dataframe** datatype
- ▶ **pandas** operations for reshaping, grouping, and more

READ the files marked READ.

READ the slides.

For these topics, it is very worthwhile to also read/watch the supplementary material.

And, most of all, do the lab sessions!