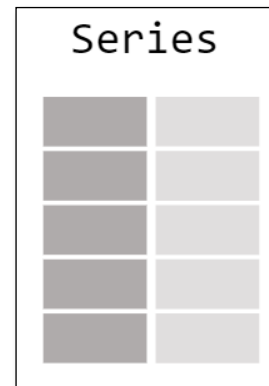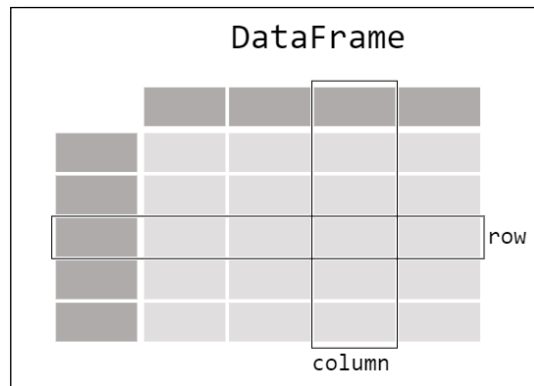# Working with Pandas-I

Chaitanya K S
Deputy Director
National Water Academy
Central Water Commission, Pune

Why Pandas?

- Pandas is a fast, powerful, flexible and easy to use open-source data analysis and manipulation tool, built on top of the Python programming language. https://pandas.pydata.org/

- Makes working with 'relational' or 'labeled' data both easy and intuitive

- Two primary data structures - Series(1-D) and DataFrame(2-D) . Each Column in a DataFrame is a Series

Why Pandas? (contd.)

- Easy handling of missing data
- Size mutability: columns can be inserted and deleted
- Automatic and explicit data alignment
- group by functionality for aggregating and transforming data
- Differently indexed data in other Python data structures can be easily converted into DataFrame objects
- Intelligent label-based slicing, fancy indexing, and subsetting of large data sets
- Intuitive merging and joining data sets
- Flexible reshaping and pivoting of data sets
- Hierarchical labelling of axes
- Robust IO tools for loading data from various formats
- Time series-specific functionality

For data scientists, working with data is typically divided into multiple stages: munging and cleaning data, analyzing / modeling it, then organizing the results of the analysis into a form suitable for plotting or tabular display. pandas is the ideal tool for all of these tasks. https://pandas.pydata.org/docs/getting_started/overview.html

Thank you