Week 3: Pandas

Kevin Lee

Department of Statistics
Western Michigan University

Introduction to Pandas

- Pandas was created by Wes McKinney in 2008.

- The name is derived from the term "panel data", an econometrics term for data sets that include observations over multiple time periods for the same individuals.

- Pandas is a newer package built on top of NumPy, and provides an efficient implementation of a `DataFrame`.

- `DataFrame`s are essentially multidimensional arrays with attached row and column labels, and often with heterogeneous types and/or missing data.

Pandas Objects

- Pandas objects can be thought of as enhanced versions of NumPy arrays in which the rows and columns are identified with labels rather than simple integer indices.

- Two fundamental Pandas data structures are `Series` and `DataFrame`.