

Pandas: DataFrame Merging

Akitaka Matsuo
Essex IADS



Merging Motivation

- There are countless situation when you want to combine two or more datasets
- Example:
 - Main dataset: Election results at constituency level
 - We have other datasets at the same level
 - Demography
 - Previous election results
 - Economic conditions
 - If we have a unique identifier, we can combine two datasets
 - Name? (not ideal, e.g. Newcastle-upon-Tyne)
 - ID code? (ideal)



Merging

- In Pandas, there are two methods for
 1. `df.merge()`: Combining two datasets based on some ID variables
 - Simpler (as we don't have to set index)
 2. `df.join()`: Combining two datasets based on index
 - Tedious but much faster once we set the index
- We will use mainly 1. in the demo



Merging