

International Data Preprocessing

Shriraj Gandhi and Emma Carlson



Motivation

Problems

- Lack of current packages focusing on international data preprocessing
- Conducting calculations “by hand” is too tedious
- Data is spread across many sources

Goals

- Provide users an all-in-one tool for this type of preprocessing
- Create a lightweight package that lends itself to ease of use and extensibility
- Make it simple for users to integrate the tool in existing workflows

Functionality

Per Capita Calculations

- Normalize numerical data using per capita calculations

Currency Inflation Adjustments

- Adjust currency data to display adjusted monetary value using inflation rates

Currency Foreign Exchange Rates

- Adjust currency data to reflect adjusted monetary value using exchange rates

Design

- Started from the top
 - o What functionality do we want to provide?
 - o What do we need to do to enable that functionality?
- Created a standard Country and Currency representation
 - o But, users dislike using custom data types in pandas dataframes
- Provided parsers to read and write from/to basic data types



Github Repository



Demo

Extensibility

Versioning Capabilities

- Promotes extensibility with external csv file that is easily updated
- Calculation metrics can be altered or added over time

Future Additions

- Time Zone adjustments
- Performance improvements
- Bundle top-level operations and parsing

Thank You!

