

Topic Outline: Shaping Data

Revised: April 6, 2016

Materials

- Today's handout: this outline
- Posted on *Topic outlines & links* page of website

Pandas 3: Data shaping

- Setup
 - Download IPython notebook
https://github.com/DaveBackus/Data_Bootcamp/blob/master/Code/IPython/bootcamp_pandas-clean.ipynb
and save Raw file in your `Data_Bootcamp` directory
In short: GitHub \Rightarrow Code \Rightarrow IPython \Rightarrow `bootcamp_pandas-clean.ipynb` \Rightarrow Raw
 - Open in Jupyter
- **Want operator**
 - Keep in mind what we want, then figure out how to get there
 - Problems: numbers contain dollar signs or commas, interpreted as strings; missing values not marked; dataset contains a lot of stuff we don't want; rows and columns are flipped.
- String methods
 - Fixing oddities in data: dollar signs, commas, etc.
- Missing values
 - Identify missing values, then Pandas will automatically work around them
- Selecting variables (columns) and observations (rows)
 - Various things we don't use much
- Boolean selection
 - Things we use — a lot
 - Comparisons, boolean selection
 - The `isin` and `contains` methods

After class

- Required
 - Submit Revised Project Ideas
- Recommended
 - Skim Project Guide
 - Skim Project Examples
 - Bounce around project ideas with friends