

QSS20: Modern Statistical Computing

Unit 07: Catchup: Merging data & LaTeX

Goals for today

- ▶ Upcoming deadlines
- ▶ Recap of reshaping data
- ▶ Merging data
- ▶ LaTeX

Goals for today

- ▶ **Upcoming deadlines**
- ▶ Recap of reshaping data
- ▶ Merging data
- ▶ LaTeX

Upcoming deadlines

DataCamp:

- ▶ **Today:** Chapters on merging data
- ▶ **Wednesday 02/01:** Chapter on regular expressions for matching
- ▶ **Optional:** Chapter on intro to HTML (essential for web-scraping)

Problem sets:

- ▶ **Problem set 1 makeup:** due Wednesday 02/01 by 11:59 PM; replaces pset 1 problem score, *or* chance for extra credit
- ▶ **Problem set 3** (reshaping, data merging, regex): due next Wednesday 02/08
- ▶ **Final project milestone 1** (research question): due next Sunday 02/12

Survey reminders (past due; please do TONIGHT):

- ▶ **Pset 2 feedback:** 21 have completed
- ▶ **Final project group preferences:** 18 have completed
- ▶ Can find both surveys [on course website](#) “Useful Links” page

Goals for today

- ▶ Upcoming deadlines
- ▶ **Recap of reshaping data**
- ▶ Merging data
- ▶ LaTeX

Recap of reshaping data

What do you remember?

Recap of reshaping data

Tips:

- ▶ *Long format* data: think of a chair viewed from front (longer than wide)
- ▶ *Wide format* data: think of a chair knocked over (wider than long)
- ▶ Inputs can seem confusing, but only three:
 - ▶ Unit of analysis (`index` or `id_vars`)—the thing we're tracking
 - ▶ Unit of variation (`columns` for `pd.pivot`)—what varies, e.g. time, distance (AKA independent variable)
 - ▶ Measurements (`values` for `pd.pivot`)—the concrete “stuff” we care about (AKA dependent variable)

Useful commands:

```
pd.pivot(longdf,  
         index='Student',  
         columns='year',  
         values=['gpa', 'ncourses'])  
  
pd.melt(widedf,  
        id_vars='Student')
```

Where we are

- ▶ Upcoming deadlines
- ▶ Recap of reshaping data
- ▶ **Merging data**
- ▶ LaTeX

Merging data (from previous class)

Slides

Merging part of [06_qss20_w23_mergereshape.pdf](#)

Activity

Merging part (section 2) of [03_resaping_merging_blank.ipynb](#)

Where we are

- ▶ Upcoming deadlines
- ▶ Recap of reshaping data
- ▶ Merging data
- ▶ **LaTeX**

LaTeX (from class on functions)

Slides

LaTeX part of [04_qss20_w23_userdefinedfunctions_latex.pdf](#)

Activity

- ▶ [Link to Overleaf template to copy over](#)
 - ▶ Click 'Menu' in top-left then Actions/'Copy Project'
- ▶ Python activity: [02_latex_output_examples_blank.ipynb](#)

We covered

- ▶ Recap of reshaping data
- ▶ Merging data
- ▶ LaTeX

Last thing: Your feedback for today

What's one thing you learned and one challenge/lingering question you encountered?