Kickoff

Peter Ganong and Maggie Shi

October 16, 2024

Ahead of today's lecture

- If you haven't already: `pip install palmerpenguins`

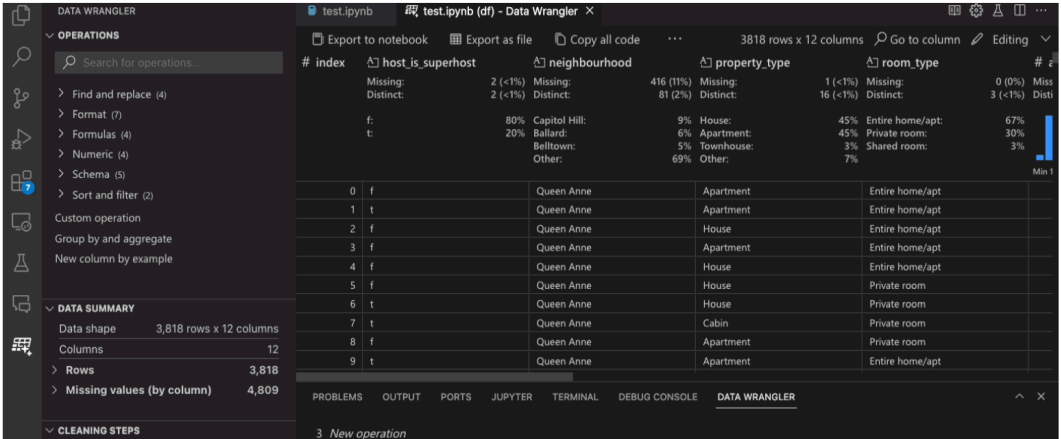
New student repo

- New student repo: https://github.com/uchicago-harris-dap/student30538 (link)
- You will have to fork it again (see instructions in README.md)
- Simpler structure: just one branch, with separate folders for `before_lecture` and `after_lecture`

Tip from Ed

- Jakub found a cool extension for viewing and cleaning data in VSCode (similar to RStudio)

- https://code.visualstudio.com/docs/datascience/data-wrangler (link)



MUD concerns about workload

- We removed a section from PS2. Revised version is in student repo.
- We are adding office hours this week
  - Akbar over zoom on Friday
  - Joaquin in person Thursday
  - *You will get priority if you post your question to the office hours thread on Ed before you come*
- Looking forward, PS3 is shorter than PS1 or PS2 as best as we can tell

MUD other

- Will we require that you transform your dataset in `altair` or `pandas`? *Answer: no, use what works best for you*
- Can `altair` functions (e.g., `transform_filter()`) be used outside of `altair`? *Answer: no, altair only transforms within the plot it's making*
- Final project: why just one member present? *Answer: presentation slots are 8 minutes and we want to reduce transition time.*
- Challenges debugging – we will do a mini-lesson on this next week
- Good questions we will have answers to by Monday (but don't have yet, sorry!)
  - What is a concrete example of a time/setting when winsorizing would be the best path for handling outliers
  - Adding better axis labels (motivated by the `diamonds` histogram for `carat`)

Final Project Optional Meetings

- Preferred but not required: send questions in advance via private Ed post tagged "Final project Ganong"
- Ganong's sections: Next week *during* lab in room 0049. Will have a signup at the start of lab.
- Shi's sections: office hours.