Kickoff

Peter Ganong and Maggie Shi

October 21, 2024

MUD

▶ *When to winsorize?* When you have an observation which has reasonable values for almost all variables, but is an extreme outlier for just one variable.

  ▶ Regression: you want to include that observation but don't want the estimates to be affected by the outlier

    ▶ If you set that cell to NA, the entire observation drops out of the regression.

  ▶ Summary statistics: you want to report the same number of observations across all your descriptive statistics

    ▶ If you set that cell to NA, your summary statistic for that variable will have fewer observations than for the other variables.

  ▶ Often used for variables that extreme extreme skewness: healthcare costs

Curriculum

▶ Finishing up visualization in the first half of today

▶ Remaining 1.5 lectures this week are on version control

Problem Sets

▶ Two problem sets released later today, will send a Canvas announcement when they are live.

▶ PS3: Git – due 10/26

    ▶ Half solo, half paired

    ▶ Should be shorter than PS1 and PS2

▶ PS4: Spatial – due **11/2**

    ▶ Paired, but longer

    ▶ First 2 out of 5 sections is just data cleaning – highly recommend you get started this week

Final Project meetings

Send questions in advance in Ed via tag "Final project ganong"

First come first serve sign up **here**. Meet in room 0049.