

Kickoff

Peter Ganong and Maggie Shi

October 21, 2024

Problem Sets

- ▶ Two problem sets released today
- ▶ 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

- ▶ *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: income, stock returns, real estate prices, healthcare costs