

# Predicting Housing Prices Presentation

## Dataset Introduction (Lauren / Devansh)

- Where the dataset came from
- Some examples of predictors and the size of the dataset
- From a Kaggle competition
  - Root mean squared logarithmic error
- A list of all of the models we tried

## Data Preprocessing (Kevin / Weston)

- Missing data and how it was dealt with
- Feature selection and how we removed unnecessary predictors
  - State specific reasons why we did so
- Dummy coding and how we accomplished that (Kevin)
- Data standardizing
- Combining a larger set of some predictors into a smaller set
  - ⊖ ~~BsmtSF1 and BsmtSF2 variables into BsmtFinSF~~
    - Apparently we didn't end up doing that
  - Exterior1st/Exterior2nd variables
  - State specific reasons why we did so

## Models (Everyone)

- **Order of presenting:** Lauren (PLS, PCR, and Forward/Backward Selection), John (Lasso and Ridge), Kevin (Regression Tree), Weston (KNN, GAM), Weston & Devansh (Bagging, Boosting, and Random Forests)
- Things to talk about
  - Kaggle scores
  - Tuning parameters
  - Conclusions about your models
  - Problems you ran into
  - Anything additional related to your specific model

## Conclusions (Westen / John)

- Explain why the tree models worked so well against the other models
  - Lots of predictors means that our models need to choose the best predictors or they will be overly complex
  - Feature selection done by the models helps a lot in decreasing test variance

- How our models fared pretty well against the other models on Kaggle with respect to the RMSLE scores (Kaggle score)
- How would you explain the results to an interested party