Slide 6:

* Cross-Validation Curve
  + LOOCV calculated for decision trees of varying sizes.
  + LOOCV involves iteratively separating observations to measure model performance on unseen data.
  + We addressed model complexity by removing less significant predictors from our dataset, ensuring our model isn't picking up noise.
  + Interestingly, our tree improved as we increased its depth.
  + We didn't fit a larger tree initially, but we'll explore this in our project report.

Slide 7:

* Bagging – Bootstrapping and Aggregating
  + Traditional decision trees use the entire dataset for predictions.
  + Bagging involves bootstrapping, using random subsets (with replacement) for each tree, reducing correlation.
  + This led to a significant reduction in MSE compared to a single tree.
  + Aggregating via weighted averaging leverages the diversity of models for accuracy and robustness.

Slide 8:

* Random Forests
  + Similar to bagging but restricts each tree to consider only a subset of predictors.
  + In our case, m = 3 for the random forest model.

Slide 9:

* Boosting!
  + Another ensemble method building multiple decision trees, but differs in tree construction and combination.
  + Uses "weak learners" or "decision stumps" and least squares loss function for regression.
  + Slow learner, recalibrating at each step.

Slide 10:

* Method Comparisons
  + RMSE, a square root of MSE, indicates model accuracy.
  + Consider the bias-variance trade-off.
  + Our data exhibits a strong linear relationship but high variability, affecting performance.
  + Linear models perform well due to their high bias and low variance, aligning with the data's linearity.
  + Boosting outperforms other ensemble methods due to its adaptive nature.
  + Variability in results highlights data variance, especially for smaller colleges.
  + Future steps include normalizing the response variable for better predictions.

Slide 11:

* Questions?