Team 9

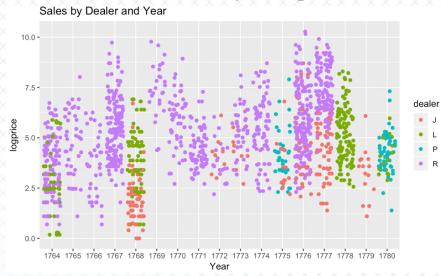*Brian T. Cozzi Zhaolin Ying Wei Zhang*

Summary of Data

Key Takeaways

These predictors fit into several relatively well-defined categories:

- Artist Characteristics

- Auction Characteristics

- Painting Content and Physical Features

Sparsely populated categories can easily lead to overfitting

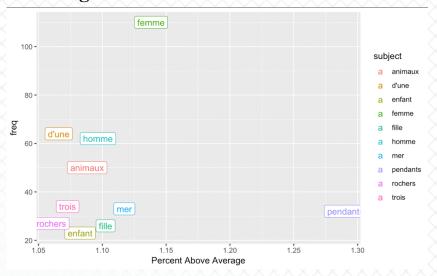Clean Data, Messy Interpretation

Text Analysis & Subject

Summary

- Split into individual words

  le   chat   est   petit

- Remove stop words

  ~~le~~   chat   ~~est~~   petit

- Remove words that are not present in at least .5% of subjects

  ~~le~~   ~~chat~~   ~~est~~   petit

89 Features Remained

High-Cost Words

Duke UNIVERSITY
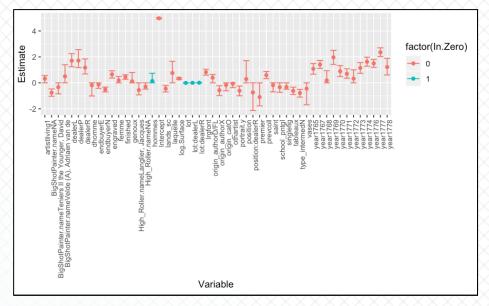
Model Choice

1. **BART**
2. **XGBoost**
3. **Random Forest**

Same performance on holdout as BMA

4. **BMA**

VARIABLES WITH POSTERIOR INCLUSION PROBABILITIES OVER .2

Top 3 Paintings