

# Team 4: The Stat Pack™

Quinn Frank

Pierre Gardan

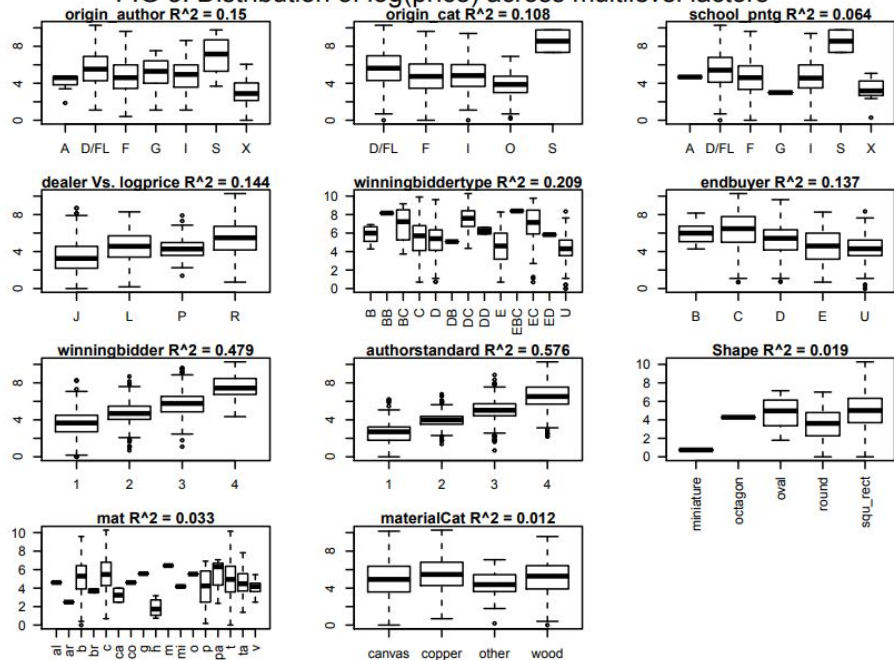
Dror Shvadron

Yiping Song



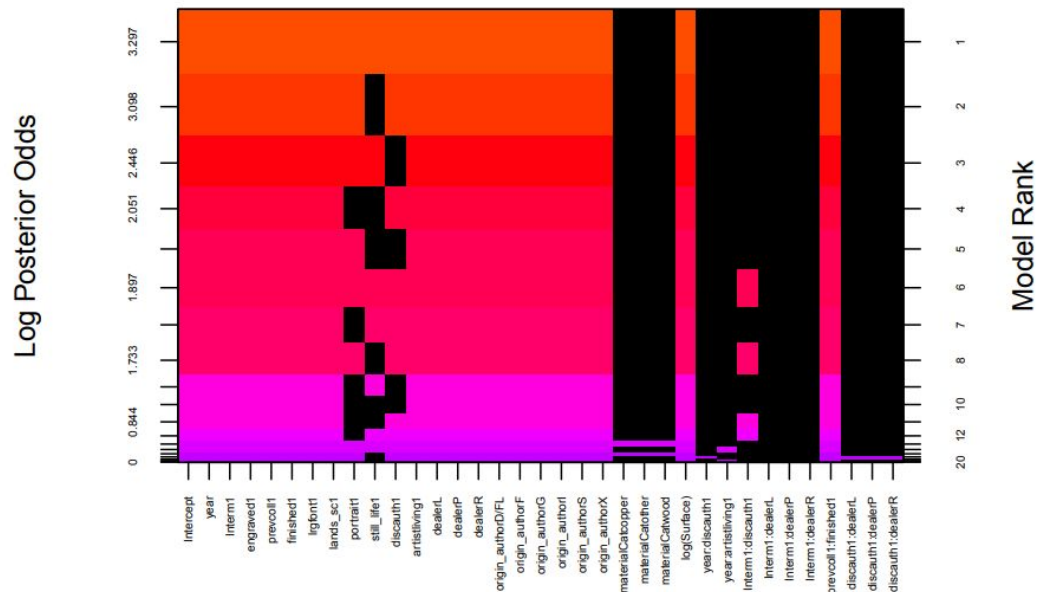
# Summary of Data

FIG 3: Distribution of log(price) across multilevel factors



- Many variables
  - Categorical (binary/multi-level)
  - Numerical
  - Textual
- Context
  - Artist
  - Sale
  - Painting features
- **Variable selection**

# Modelling Decisions



- Manual selection of variables
- AIC stepwise selection to determine significant interactions
- BMA for more shrinkage/selection
- k-fold cross-validation for selection of final model

## Questions:

- Should Year be numeric or a factor?
- Use model averaging or just HPM?
- Enumeration or MCMC?
- Dimension of search space?
- Which diagnostics to focus on?

# Results of Model

- Some significant predictors: *year, log(Surface), dealer, origin\_author, Interm, lrgfont, prevcoll:finished*
  - marginal inclusion probability = 1
  - only one interaction remained in HPM

## Cross validation:

- Similar in-sample/out-of-sample RMSE
- Comparatively low variance
- Leaderboard (test set):  
**1282.47** RMSE, **0.957** coverage

Table 5: k-fold CV for final model (HPM estimator)

	null_train	train_rmse	train_covg	null_test	test_rmse	test_covg
	2133.484	1421.483	0.946	2627.819	1948.392	0.933
	2154.045	1519.791	0.952	2557.785	1697.960	0.930
	2344.014	1578.597	0.948	1769.118	1310.456	0.959
	2182.964	1569.932	0.956	2457.883	1510.672	0.923
	2377.717	1612.914	0.951	1576.303	1277.595	0.967
average	2238.445	1540.543	0.951	2197.781	1549.015	0.942

# Top Paintings

1. Karel du Jardin, *Un Marchand d'ovietan, en habit de Scaramouche, monte sur son theatre...* (**predicted 17,765.73 livres**)

- *Le trésor de la curiosité* by Charles Blanc (1857)  
claims price is 17,720 livres

2. Rembrandt, *The Night Watch* (top) (**predicted 13,639.27 livres**)

3. Peter Paul Rubens, *Adoration of the Magi* (bottom) (**predicted 13,572.85 livres**)



# Insights

- Linear modeling does well in predicting painting prices
  - Can't rely only on automatic procedures for model selection
  - BMA can average models, shrinks coefficients to avoid overfitting
  - If one model dominates, BMA provides easy selection procedure
- 
- Factors that drove painting prices:
    - **Sale details** (Who bought it? From whom? When?)
    - **Who painted it** (What country? Still living?)
    - **Provenance** (history of the painting/previous owners)
    - Not so much the actual content of the painting