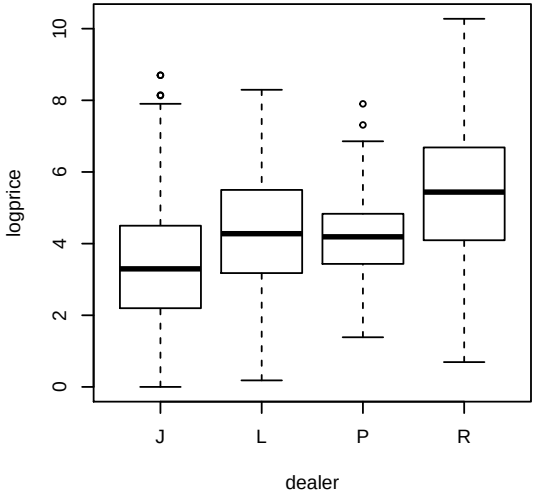
STA 521 Final Project

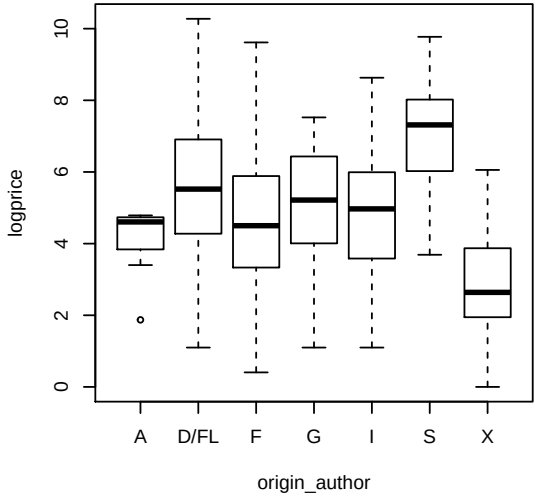
Team 10: Bin Han, Jingyi Zhang, Jonathan Klus

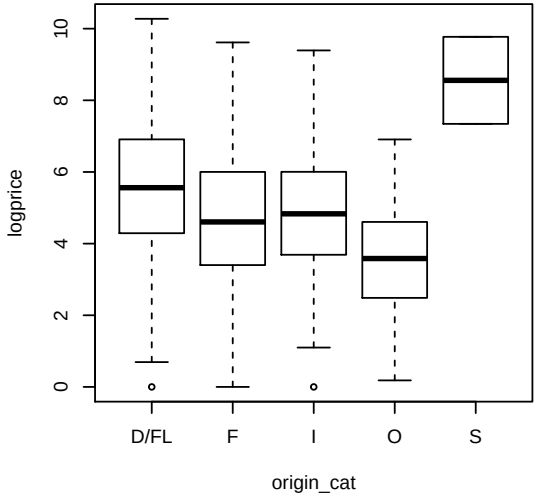
12 December 2018

EDA & Data Cleaning

- Removed variables: `lot`, `sale`, `price`, `count`, `subject`, `authorstandard`, `winningbidder`, `Surface_Rnd`, `Surface_Rect`, `material`, `mat`
- Recoded many categorical variables, for example:
  a. `endbuyer`: "n/a" & "" – "X"
  b. `authorstyle`: "n/a" & "" – 0; others: 1
  c. `materialCat`: "n/a" & "" – "other"







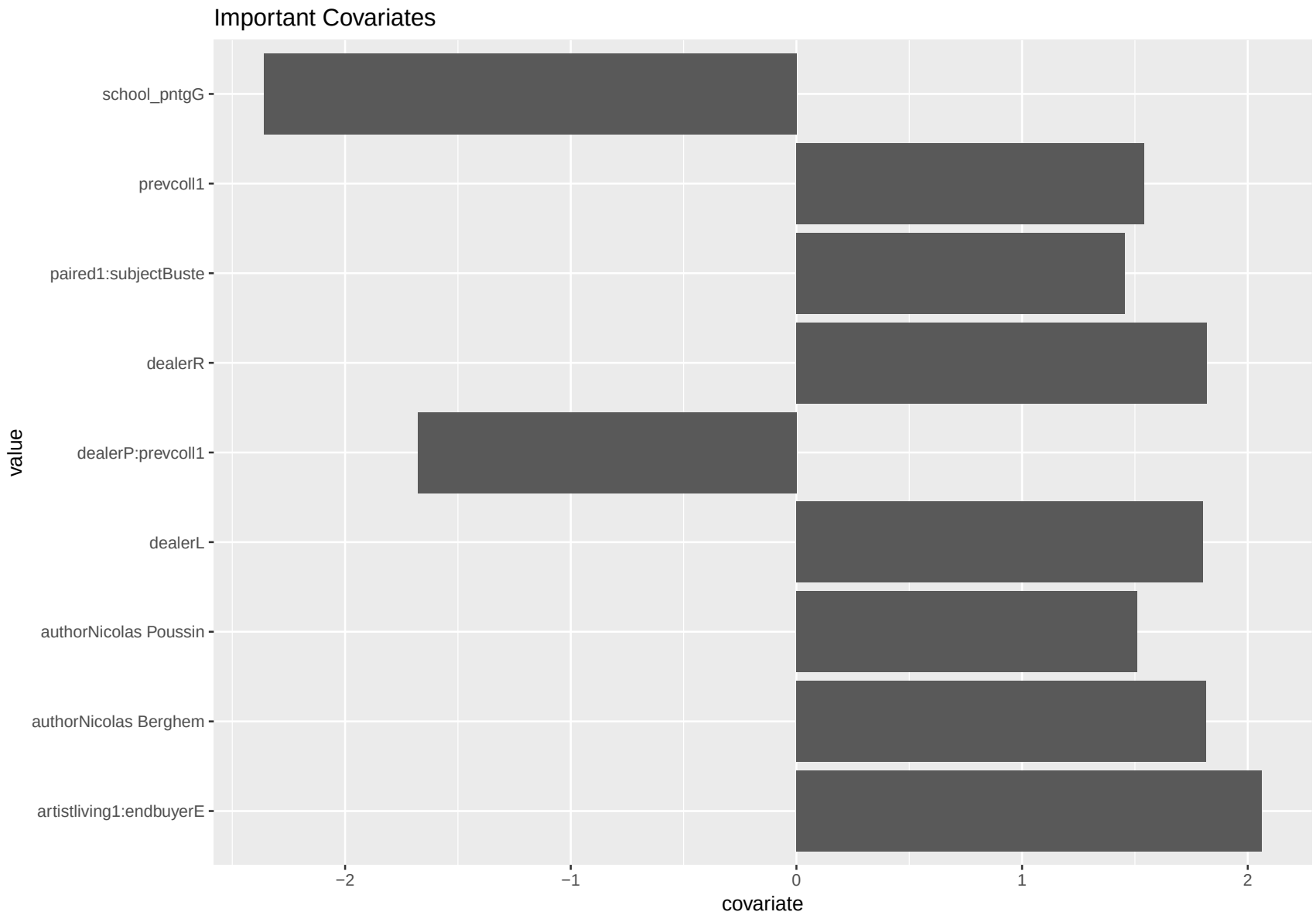
BMA + AIC

author	sum
David Teniers	46
Philippe Wouvermans	27
Francois Boucher	26

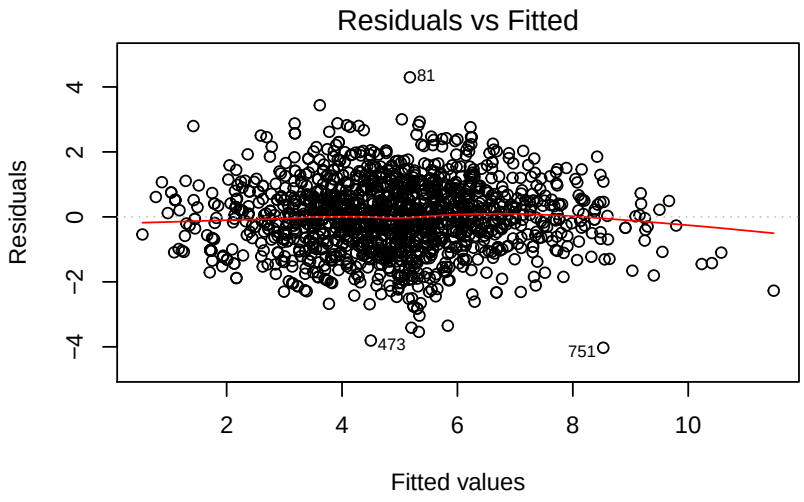
subject	sum
Paysage	324
People	194
Saint	121

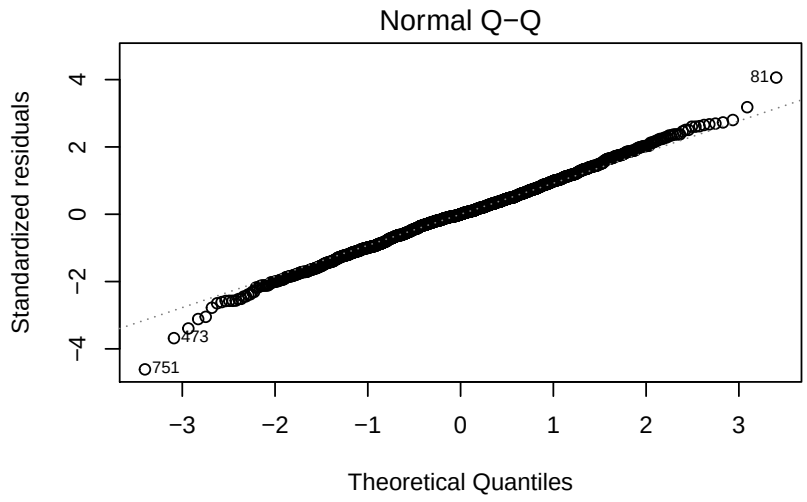
- BMA: used BPM in BMA to choose base variables that have higher posterior density and good for prediction.
- AIC: generated all possible interactions and applied AIC to choose important features.
- Final model: 19 base vairables; 24 interactions;

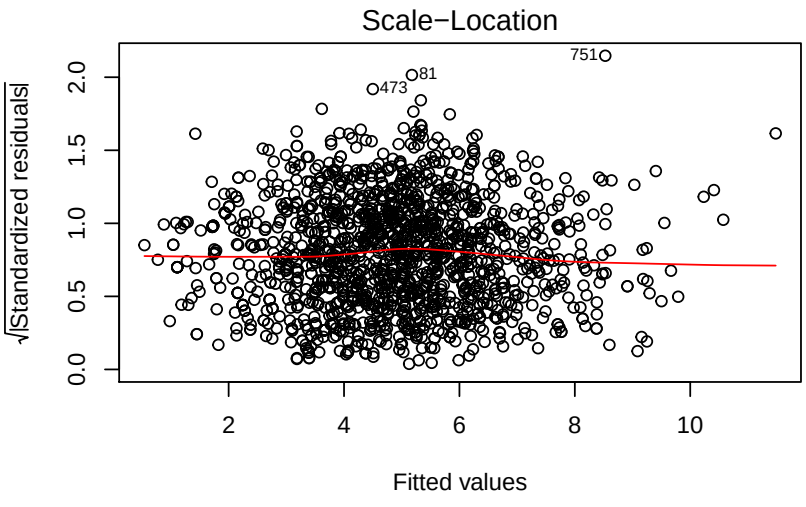
Important Features

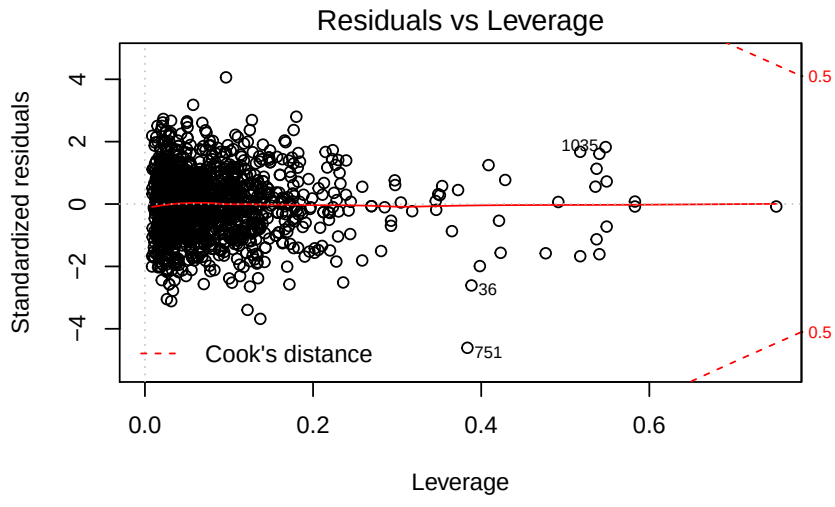


Model Diagnostics









Interesting Finding

- The most expensive five paintings predicted from the validation set all come from the same artist: Nicolaes Pieterszoon Berchem;
- Most of the more expensive paintings are auctioned from the dealer with initial "R";
- The endbuyer of these expensive paintings are, instead of a buyer or a dealer who may serve as an intermediary, collectors themselves;
- Extremely large surface are doesn't mean high price;
- This painting is an example of highly priced paintings:

1. https://upload.wikimedia.org/wikipedia/commons/e/e0/David_Teniers_the_Younger_-_The_Feast_of_the_Prodigal_Son.jpg)
2. https://upload.wikimedia.org/wikipedia/commons/e/e8/Berchem%2C_Nicolaes_~_Landscape_with_Herdsmen_Gathering_Sticks%2C_early_1650s%2C_oil_on_panel%2C_private_collection%2C_New_York.jpg

Artist

Painting 1

Figure 2: Pieterszoon's Painting

Citation:

1. https://en.wikipedia.org/wiki/Nicolaes_Pieterszoon_Berchem#/media/File:Portrait_of_Nicolaes_Berchem_by_Jan_Stolker.jpg

Painting 2