

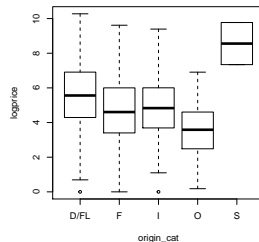
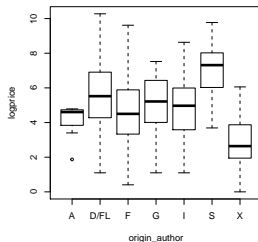
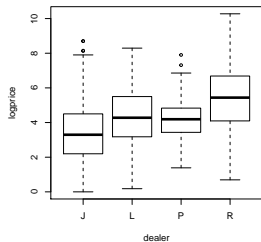
STA 521 Final Project

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:
 - endbuyer: "n/a" & "" – "X"
 - authorstyle: "n/a" & "" – 0; others: 1
 - materialCat: "n/a" & "" – "other"



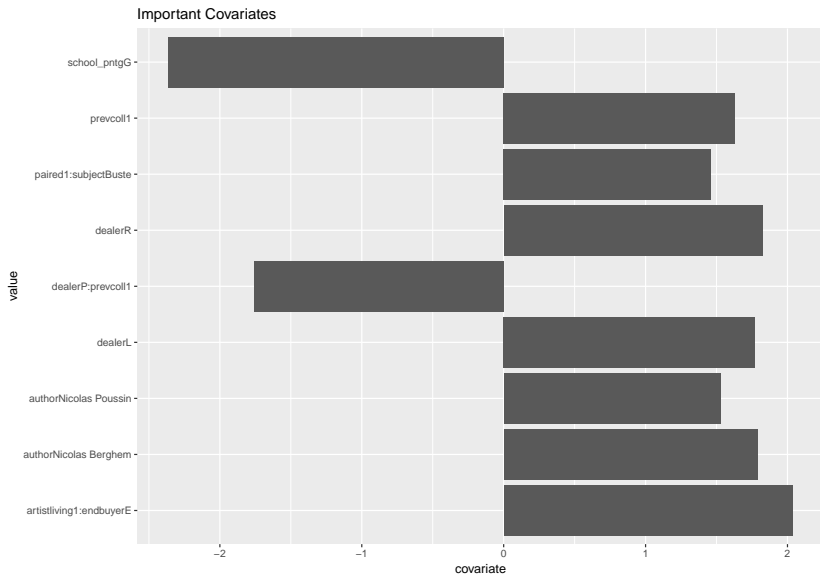
BMA + AIC

author	sum
David Teniers	46
Philippe Wouvermans	27
Francois Boucher	26

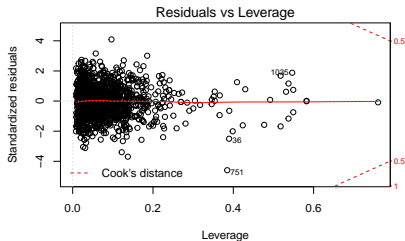
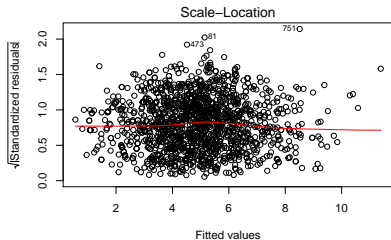
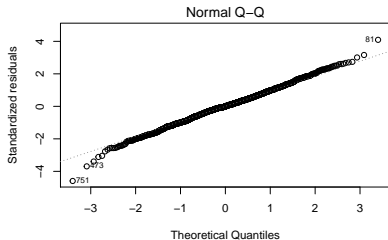
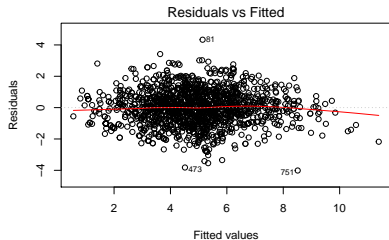
subject	sum
Paysage	324
People	194
Saint	121

- ▶ 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 variables; 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 deal 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 area doesn't mean high price;
- ▶ This painting is an example of highly priced paintings:

Artist



Painting 1



Figure 2: Pieterszoon's Painting

Painting 2

