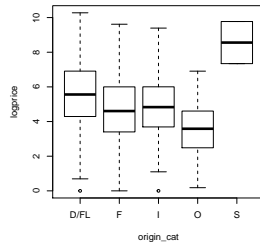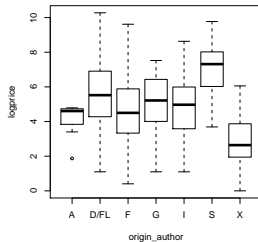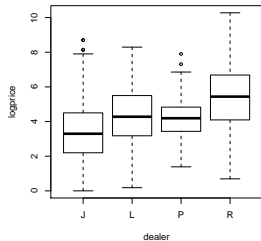# STA 521 Final Project

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"
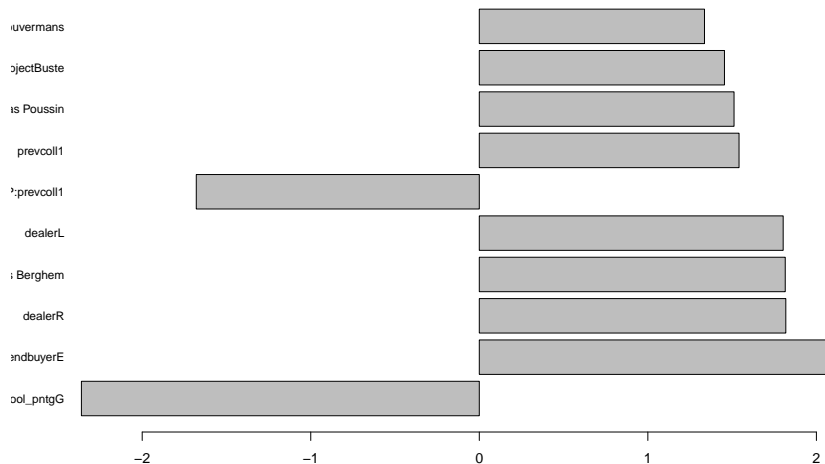 b. `authorstyle`: "n/a" & "" – 0; others: 1
  c. `materialCat`: "n/a" & "" – "other"

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