

Part-I-Writeup

To get started read in the training data:

1. Introduction:

In this study, we are looking at the auction prices of paintings in 18th century Paris. Specifically, through the assistance of model built based on existing training data, we wish to understand the factors that drive the prices of the paintings, and then be able to predict auction prices based on characteristics of a certain painting. After fitting appropriate model, we also intend to detect specific paintings that are either underpriced or overpriced based on the selected model.

One of the main task and challenge is to narrow down the number of potential predictors from 59 to less than 20 while maintaining a high performance of the model. But being able to explain the results and provide some recommendations to individuals without statistical background is equally important and challenging. Therefore, we aim at balancing the performance of model prediction, closeness to true model, simplicity, and interpretability.

2. Exploratory data analysis:

A) Data summary & cleaning

To start with, we looked at the summary of the original trainig data. There are few numeric variables and a lot of binary variables. Some variables, such as **Interm**, **Surface**, **Height_in** etc. have missing values, which need to be taken care of. The followings steps are how we cleaned the data:

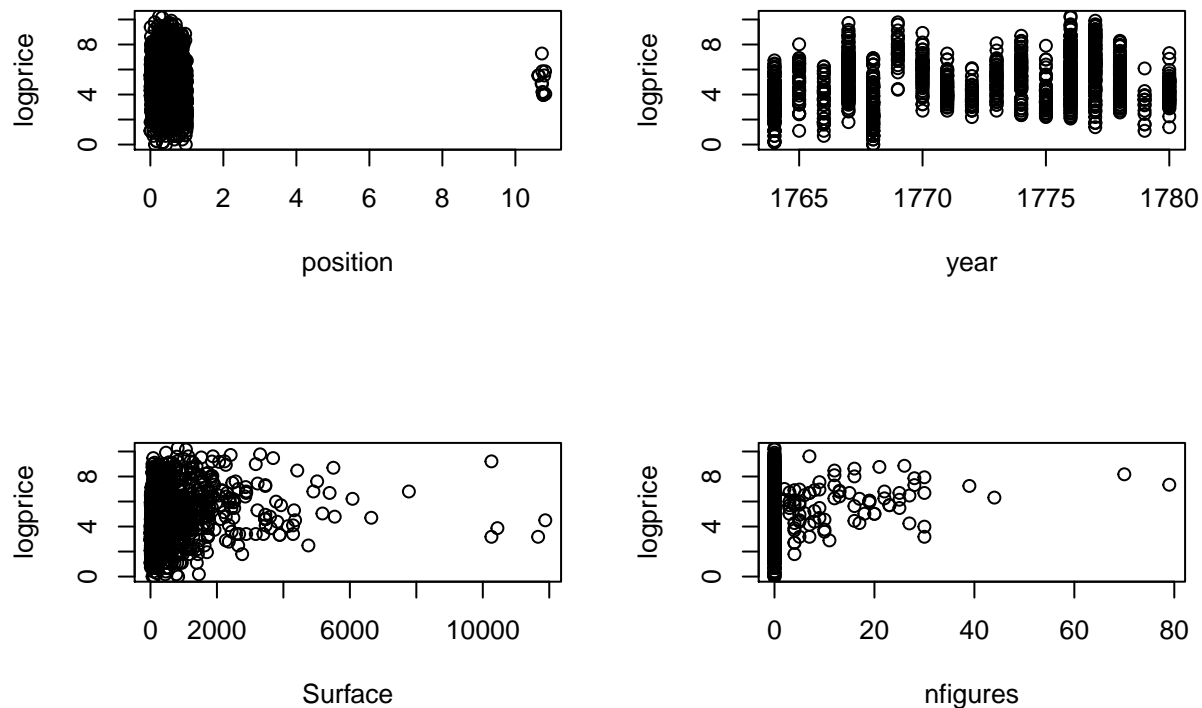
- a. The first step we did was to get rid of intuitively useless variables, including: **lot**, **sale**, **price**, **count**, **subject**, **authorstandard**, **author**, **winningbidder**, and 'other'. The are not useful in predicting the response variable (such as names)
- b. By further screening the variables, we found out that **Surface** and **Surface_Rnd**, **Surface_Rect** are corerlated, which are based on the value of **Height_in**, **Width_in**, and **Diam_in**. We decided to use **Surface** in our initial model. The same issue happened to **material**, **mat**, and **materialCat**. The latter one recodes the previous one. Therefore, we used **materialCat**. We applied the same strategy to keep **landsALL** and get rid of other variables related with landscape.
- c. For those variables that have multiple levels, to be consistent with how the data was originally coded, we recoded the missing levels as "X", which stands for "no information". For **materialCat** and **Shape**, since there are so many levels, we grouped some levels with few observations together, coded as "other" group. The rest binary vairables are changed into factor.
- d. Then we dealt with the missing values in **Surface** and **Interm**. We used the package "mice" to address this problem, which uses the observed values in the dataset to impute the missing values. It prevents directly throwing away the missing values, which results in lossing a large amont of information for prediction.

```
##
## iter imp variable
## 1 1 Interm* Surface*
## 1 2 Interm Surface*
## 1 3 Interm* Surface*
## 1 4 Interm* Surface*
## 1 5 Interm* Surface*
```

```
## 2 1 Interm* Surface*
## 2 2 Interm* Surface*
## 2 3 Interm* Surface*
## 2 4 Interm* Surface*
## 2 5 Interm* Surface*
## 3 1 Interm* Surface*
## 3 2 Interm* Surface*
## 3 3 Interm* Surface*
## 3 4 Interm* Surface*
## 3 5 Interm* Surface*
## 4 1 Interm* Surface*
## 4 2 Interm* Surface*
## 4 3 Interm* Surface*
## 4 4 Interm* Surface*
## 4 5 Interm Surface*
## 5 1 Interm* Surface*
## 5 2 Interm* Surface*
## 5 3 Interm* Surface*
## 5 4 Interm* Surface*
## 5 5 Interm* Surface*
## * Please inspect the loggedEvents
```

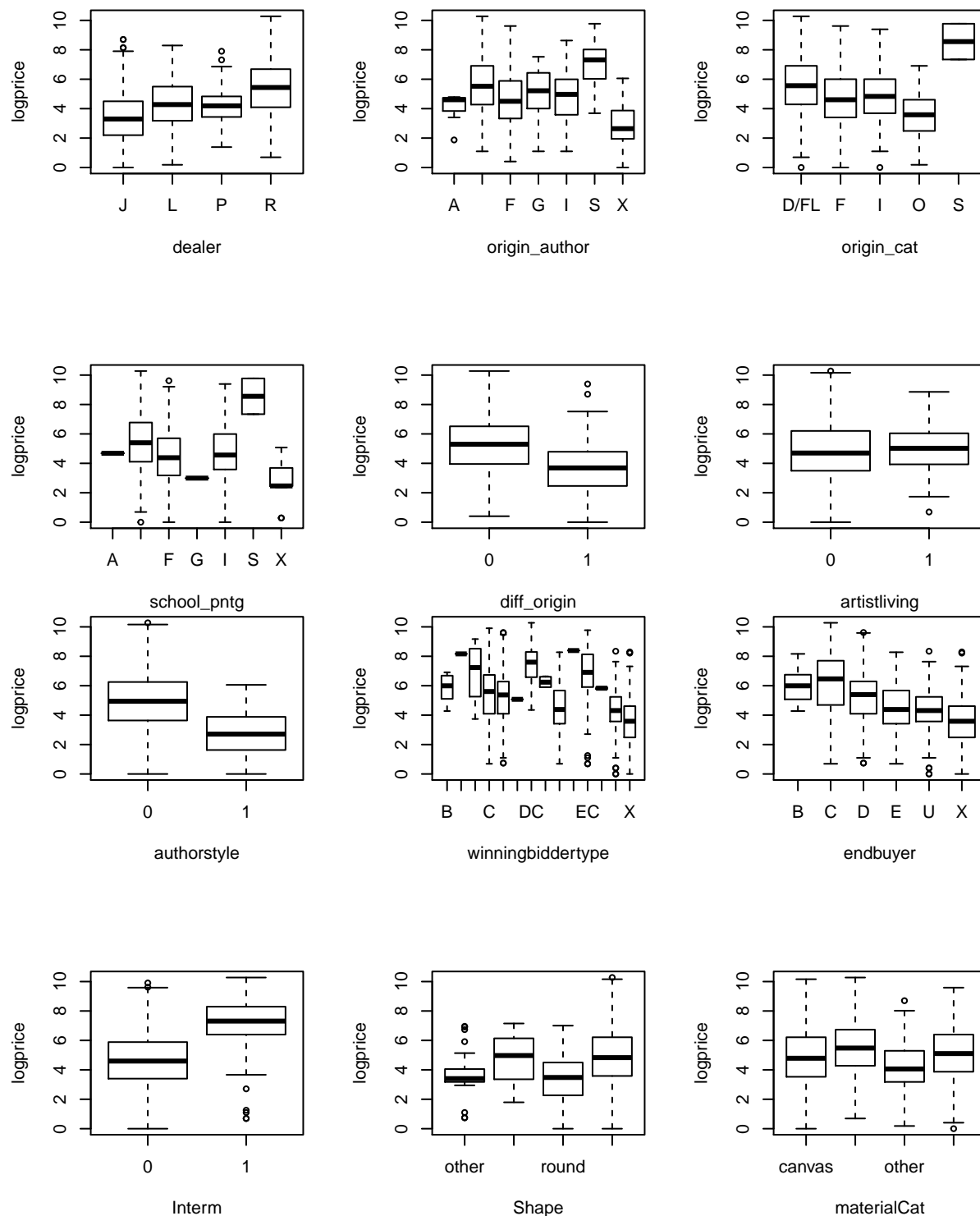
B). Plots

Then we analyzed the relationship between those left features and the response variable. With the scatter plots, we can roughly determine which variables can be put into the initial model. For categorical variables, we want to check if the `logprice` spans different ranges in different levels. For numeric variables, we want to check if there is a clear relationship between them and `logprice`.



For numeric variables, we see that `Surface` and `nfigures` seem to show some weak but positive relationship

with `logprice`. Since there are several extremely large values in `position` (potentially outliers), it is hard to see that real pattern between the majority of points and `logprice`. But we'll keep it in the model first.



Since there are 33 categorical variables, we don't show the boxplots for all of them. But applied the same method to check all the categorical variables. The following variables show some differences in `logprice` at different levels (not considering the magnitude of the difference at this time): `dealer`, `origin_author`, `origin_cat`, `school_pntg`, `diff_origin`, `authorstyle`, `endbuyer`, `Interm`, `Shape`, `materialCat`, `engraved`,

prevcoll, figures, finished, lrgfont, othgenre, discauth, and still_life.

If we were to choose 10 best predictive variables for predicting, we would consider the magnitude of differences and the strength of relationships. The 10 variables we choose are: Surface, dealer, school_pntg, diff_origin, authorstyle, endbuyer, Interm, prevcoll, engraved, lrgfont.

3. Development and assessment of an initial model

Initial Model

## [1] "Intercept"	"dealerL"	"dealerR"
## [4] "year"	"school_pntgD/FL"	"school_pntgI"
## [7] "diff_origin1"	"artistliving1"	"authorstyle1"
## [10] "winningbiddertypeDC"	"winningbiddertypeE"	"winningbiddertypeEBC"
## [13] "endbuyerD"	"endbuyerU"	"endbuyerX"
## [16] "Interm1"	"Shaperound"	"Surface"
## [19] "materialCatother"	"nfigures"	"engraved1"
## [22] "prevcoll1"	"paired1"	"finished1"
## [25] "lrgfont1"	"peasant1"	"still_life1"
## [28] "discauth1"		

## [1] "Intercept"	"dealerL"	"dealerR"
## [4] "year"	"origin_authorF"	"school_pntgD/FL"
## [7] "school_pntgX"	"diff_origin1"	"artistliving1"
## [10] "authorstyle1"	"winningbiddertypeC"	"winningbiddertypeEBC"
## [13] "endbuyerE"	"endbuyerU"	"endbuyerX"
## [16] "Interm1"	"Shaperound"	"Surface"
## [19] "materialCatother"	"engraved1"	"prevcoll1"
## [22] "paired1"	"figures1"	"finished1"
## [25] "lrgfont1"	"peasant1"	"portrait1"
## [28] "still_life1"	"discauth1"	"pastorale1"

The EDA process gives us an initial idea of which variables to drop out to reduce the dimension, and which variables might be significant in explaining the variation in logprice. But before we built the initial model, we applied BMA, Bayesian Model Averaging, to systematically choose which base variables that have higher posterior probabilities to be in the initial model. We experimented two modelpriors, “JZS” and “g-prior”, which gave us two sets of variables listed above. Then we picked up the common ones from Best Predictive Model(BPM).

Then we fit the linear regression model using the chosen features and all their possible interactions. From the summary table, the $R^2 = 0.6928$, which is fairly high. But we realized that lots of estimated coefficients for interactions are NAs, indicating that some levels in those variables have too few observations to be estimated. Therefore, we need to further reduce the dimension through variable selection.

```
##
## Call:
## lm(formula = logprice ~ (dealer + school_pntg + diff_origin +
##   artistliving + endbuyer + authorstyle + Interm + Shape +
##   Surface + engraved + prevcoll + paired + finished + lrgfont +
##   portrait + discauth + still_life)^2, data = paintings_train_2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.9813 -0.6650  0.0151  0.6177  4.3925
##
```

```
## Coefficients: (132 not defined because of singularities)
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    7.555e+00  7.201e+00   1.049 0.294341
## dealerL        9.292e-01  2.812e+00   0.330 0.741112
## dealerP        5.612e-01  4.599e+00   0.122 0.902907
## dealerR       -1.477e+00  3.699e+00  -0.399 0.689734
## school_pntgD/FL 2.331e+00  3.764e+00   0.619 0.535839
## school_pntgF   -2.145e-01  4.622e+00  -0.046 0.962996
## school_pntgG   -1.942e+00  2.523e+00  -0.770 0.441652
## school_pntgI    5.168e-01  2.382e+00   0.217 0.828293
## school_pntgS   -4.051e+00  6.433e+00  -0.630 0.528999
## school_pntgX   -4.416e-01  1.964e+00  -0.225 0.822110
## diff_origin1   -5.289e-01  1.878e+00  -0.282 0.778313
## artistliving1  -2.381e-01  3.592e+00  -0.066 0.947171
## endbuyerC      -3.677e+00  4.797e+00  -0.767 0.443516
## endbuyerD      -3.523e+00  4.962e+00  -0.710 0.477904
## endbuyerE      -6.351e+00  4.423e+00  -1.436 0.151297
## endbuyerU      -3.494e+00  4.559e+00  -0.767 0.443517
## endbuyerX      -1.108e+00  2.282e+00  -0.486 0.627305
## authorstyle1   -1.401e+00  8.043e-01  -1.742 0.081747
## Interm1        -7.569e-01  1.766e+00  -0.429 0.668283
## Shapeoval      -2.726e+00  1.045e+01  -0.261 0.794137
## Shaperound     -5.900e+00  6.563e+00  -0.899 0.368847
## Shapesqu_rect  -4.920e+00  6.368e+00  -0.773 0.439886
## Surface        3.719e-03  3.810e-03   0.976 0.329200
## engraved1     -1.655e+00  1.016e+00  -1.630 0.103392
## prevcoll1      2.101e+00  8.786e-01   2.391 0.016952 *
## paired1        2.541e+00  1.811e+00   1.403 0.160781
## finished1      1.978e+00  2.627e+00   0.753 0.451671
## lrgfont1       4.070e-02  1.499e+00   0.027 0.978349
## portrait1     -6.428e-02  1.652e+00  -0.039 0.968971
## discauth1      2.591e+00  1.816e+00   1.427 0.153842
## still_life1   -1.976e-01  1.674e+00  -0.118 0.906061
## dealerL:school_pntgD/FL 3.765e-01  1.381e+00   0.273 0.785178
## dealerP:school_pntgD/FL -1.159e+00  1.716e+00  -0.675 0.499706
## dealerR:school_pntgD/FL 4.040e-01  3.992e-01   1.012 0.311693
## dealerL:school_pntgF 7.538e-01  1.387e+00   0.544 0.586798
## dealerP:school_pntgF -8.767e-01  1.732e+00  -0.506 0.612752
## dealerR:school_pntgF 5.403e-01  4.267e-01   1.266 0.205663
## dealerL:school_pntgG      NA      NA      NA      NA
## dealerP:school_pntgG      NA      NA      NA      NA
## dealerR:school_pntgG      NA      NA      NA      NA
## dealerL:school_pntgI 7.490e-01  1.413e+00   0.530 0.596037
## dealerP:school_pntgI -1.278e+00  1.836e+00  -0.696 0.486489
## dealerR:school_pntgI      NA      NA      NA      NA
## dealerL:school_pntgS      NA      NA      NA      NA
## dealerP:school_pntgS      NA      NA      NA      NA
## dealerR:school_pntgS      NA      NA      NA      NA
## dealerL:school_pntgX      NA      NA      NA      NA
## dealerP:school_pntgX      NA      NA      NA      NA
## dealerR:school_pntgX      NA      NA      NA      NA
## dealerL:diff_origin1 -2.994e-01  4.776e-01  -0.627 0.530822
## dealerP:diff_origin1 -7.648e-01  5.324e-01  -1.436 0.151122
## dealerR:diff_origin1 -3.449e-01  3.988e-01  -0.865 0.387256
```

## dealerL:artistliving1	-1.299e+00	4.903e-01	-2.649	0.008172	**
## dealerP:artistliving1	-8.354e-01	6.015e-01	-1.389	0.165143	
## dealerR:artistliving1	-4.306e-01	4.077e-01	-1.056	0.291146	
## dealerL:endbuyerC	1.580e+00	2.558e+00	0.618	0.536854	
## dealerP:endbuyerC	8.190e-01	9.695e-01	0.845	0.398399	
## dealerR:endbuyerC	5.511e-01	5.562e-01	0.991	0.321967	
## dealerL:endbuyerD	1.479e+00	2.488e+00	0.594	0.552468	
## dealerP:endbuyerD	9.763e-02	6.311e-01	0.155	0.877089	
## dealerR:endbuyerD	7.203e-01	3.293e-01	2.187	0.028913	*
## dealerL:endbuyerE	1.863e+00	2.552e+00	0.730	0.465554	
## dealerP:endbuyerE	6.440e-01	7.503e-01	0.858	0.390908	
## dealerR:endbuyerE	3.378e-01	4.642e-01	0.728	0.466999	
## dealerL:endbuyerU	2.287e+00	2.573e+00	0.889	0.374269	
## dealerP:endbuyerU	6.620e-01	6.667e-01	0.993	0.320920	
## dealerR:endbuyerU	6.212e-01	3.892e-01	1.596	0.110752	
## dealerL:endbuyerX	6.834e-01	2.451e+00	0.279	0.780455	
## dealerP:endbuyerX	NA	NA	NA	NA	
## dealerR:endbuyerX	NA	NA	NA	NA	
## dealerL:authorstyle1	-1.080e+00	6.410e-01	-1.685	0.092161	.
## dealerP:authorstyle1	1.577e+00	8.245e-01	1.913	0.056025	.
## dealerR:authorstyle1	5.180e-01	5.180e-01	1.000	0.317496	
## dealerL:Interm1	1.611e+00	1.199e+00	1.343	0.179399	
## dealerP:Interm1	2.003e+00	1.796e+00	1.116	0.264838	
## dealerR:Interm1	2.850e+00	7.887e-01	3.613	0.000315	***
## dealerL:Shapeoval	NA	NA	NA	NA	
## dealerP:Shapeoval	1.680e+00	1.341e+01	0.125	0.900285	
## dealerR:Shapeoval	5.320e-01	7.822e+00	0.068	0.945791	
## dealerL:Shaperound	1.075e-02	1.482e+00	0.007	0.994214	
## dealerP:Shaperound	NA	NA	NA	NA	
## dealerR:Shaperound	3.272e+00	3.772e+00	0.867	0.385887	
## dealerL:Shapesqu_rect	NA	NA	NA	NA	
## dealerP:Shapesqu_rect	1.627e+00	4.134e+00	0.394	0.693929	
## dealerR:Shapesqu_rect	2.539e+00	3.678e+00	0.690	0.490041	
## dealerL:Surface	-2.520e-04	3.257e-04	-0.774	0.439367	
## dealerP:Surface	-4.623e-04	4.113e-04	-1.124	0.261204	
## dealerR:Surface	-3.807e-04	2.638e-04	-1.443	0.149328	
## dealerL:engraved1	2.590e+00	1.200e+00	2.159	0.031079	*
## dealerP:engraved1	NA	NA	NA	NA	
## dealerR:engraved1	1.091e-01	7.314e-01	0.149	0.881480	
## dealerL:prevcoll1	-6.306e-01	9.818e-01	-0.642	0.520790	
## dealerP:prevcoll1	-6.050e-02	1.347e+00	-0.045	0.964174	
## dealerR:prevcoll1	-8.302e-01	1.009e+00	-0.823	0.410604	
## dealerL:paired1	-1.096e+00	3.414e-01	-3.211	0.001357	**
## dealerP:paired1	-6.558e-01	4.338e-01	-1.512	0.130831	
## dealerR:paired1	-6.314e-01	2.662e-01	-2.372	0.017863	*
## dealerL:finished1	-8.389e-01	5.937e-01	-1.413	0.157891	
## dealerP:finished1	-7.543e-01	5.020e-01	-1.502	0.133234	
## dealerR:finished1	-1.003e+00	3.306e-01	-3.035	0.002458	**
## dealerL:lrgfont1	NA	NA	NA	NA	
## dealerP:lrgfont1	NA	NA	NA	NA	
## dealerR:lrgfont1	NA	NA	NA	NA	
## dealerL:portrait1	-1.014e+00	1.059e+00	-0.958	0.338494	
## dealerP:portrait1	NA	NA	NA	NA	
## dealerR:portrait1	7.163e-01	7.079e-01	1.012	0.311803	

## dealerL:discauth1	7.900e-02	1.105e+00	0.072	0.943009
## dealerP:discauth1	-1.498e+00	1.391e+00	-1.077	0.281746
## dealerR:discauth1	-1.178e+00	4.709e-01	-2.502	0.012489 *
## dealerL:still_life1	2.989e-01	7.499e-01	0.399	0.690313
## dealerP:still_life1	-6.333e-01	1.484e+00	-0.427	0.669573
## dealerR:still_life1	3.063e-03	7.010e-01	0.004	0.996514
## school_pntgD/FL:diff_origin1	-2.885e-01	3.397e-01	-0.849	0.395870
## school_pntgF:diff_origin1	-2.781e-01	3.608e-01	-0.771	0.440964
## school_pntgG:diff_origin1	NA	NA	NA	NA
## school_pntgI:diff_origin1	NA	NA	NA	NA
## school_pntgS:diff_origin1	NA	NA	NA	NA
## school_pntgX:diff_origin1	NA	NA	NA	NA
## school_pntgD/FL:artistliving1	-3.860e-01	1.444e+00	-0.267	0.789308
## school_pntgF:artistliving1	3.277e-01	1.429e+00	0.229	0.818651
## school_pntgG:artistliving1	NA	NA	NA	NA
## school_pntgI:artistliving1	-9.787e-01	1.632e+00	-0.600	0.548839
## school_pntgS:artistliving1	NA	NA	NA	NA
## school_pntgX:artistliving1	NA	NA	NA	NA
## school_pntgD/FL:endbuyerC	-7.303e-01	2.786e+00	-0.262	0.793233
## school_pntgF:endbuyerC	-7.035e-01	2.674e+00	-0.263	0.792519
## school_pntgG:endbuyerC	NA	NA	NA	NA
## school_pntgI:endbuyerC	2.294e-01	1.572e+00	0.146	0.883975
## school_pntgS:endbuyerC	NA	NA	NA	NA
## school_pntgX:endbuyerC	NA	NA	NA	NA
## school_pntgD/FL:endbuyerD	-1.823e+00	2.729e+00	-0.668	0.504177
## school_pntgF:endbuyerD	-1.241e+00	2.627e+00	-0.472	0.636867
## school_pntgG:endbuyerD	NA	NA	NA	NA
## school_pntgI:endbuyerD	-6.711e-01	1.424e+00	-0.471	0.637520
## school_pntgS:endbuyerD	NA	NA	NA	NA
## school_pntgX:endbuyerD	NA	NA	NA	NA
## school_pntgD/FL:endbuyerE	-7.770e-01	2.372e+00	-0.328	0.743336
## school_pntgF:endbuyerE	-3.286e-01	2.244e+00	-0.146	0.883623
## school_pntgG:endbuyerE	NA	NA	NA	NA
## school_pntgI:endbuyerE	NA	NA	NA	NA
## school_pntgS:endbuyerE	NA	NA	NA	NA
## school_pntgX:endbuyerE	NA	NA	NA	NA
## school_pntgD/FL:endbuyerU	-5.398e-01	2.901e+00	-0.186	0.852423
## school_pntgF:endbuyerU	-5.191e-01	2.811e+00	-0.185	0.853490
## school_pntgG:endbuyerU	NA	NA	NA	NA
## school_pntgI:endbuyerU	5.016e-01	1.736e+00	0.289	0.772640
## school_pntgS:endbuyerU	NA	NA	NA	NA
## school_pntgX:endbuyerU	NA	NA	NA	NA
## school_pntgD/FL:endbuyerX	-9.805e-01	2.363e+00	-0.415	0.678286
## school_pntgF:endbuyerX	-3.860e-01	2.240e+00	-0.172	0.863203
## school_pntgG:endbuyerX	NA	NA	NA	NA
## school_pntgI:endbuyerX	NA	NA	NA	NA
## school_pntgS:endbuyerX	NA	NA	NA	NA
## school_pntgX:endbuyerX	NA	NA	NA	NA
## school_pntgD/FL:authorstyle1	-4.789e-01	4.676e-01	-1.024	0.306018
## school_pntgF:authorstyle1	-7.266e-02	5.108e-01	-0.142	0.886917
## school_pntgG:authorstyle1	NA	NA	NA	NA
## school_pntgI:authorstyle1	NA	NA	NA	NA
## school_pntgS:authorstyle1	NA	NA	NA	NA
## school_pntgX:authorstyle1	NA	NA	NA	NA

## school_pntgD/FL:Interm1	-2.403e-01	5.974e-01	-0.402	0.687541
## school_pntgF:Interm1	4.300e-01	6.574e-01	0.654	0.513148
## school_pntgG:Interm1	NA	NA	NA	NA
## school_pntgI:Interm1	NA	NA	NA	NA
## school_pntgS:Interm1	NA	NA	NA	NA
## school_pntgX:Interm1	NA	NA	NA	NA
## school_pntgD/FL:Shapeoval	-8.002e-01	2.795e+00	-0.286	0.774674
## school_pntgF:Shapeoval	-4.402e-01	7.683e+00	-0.057	0.954319
## school_pntgG:Shapeoval	NA	NA	NA	NA
## school_pntgI:Shapeoval	NA	NA	NA	NA
## school_pntgS:Shapeoval	NA	NA	NA	NA
## school_pntgX:Shapeoval	NA	NA	NA	NA
## school_pntgD/FL:Shaperound	-1.867e+00	2.506e+00	-0.745	0.456360
## school_pntgF:Shaperound	1.309e+00	3.524e+00	0.371	0.710425
## school_pntgG:Shaperound	NA	NA	NA	NA
## school_pntgI:Shaperound	NA	NA	NA	NA
## school_pntgS:Shaperound	NA	NA	NA	NA
## school_pntgX:Shaperound	NA	NA	NA	NA
## school_pntgD/FL:Shapesqu_rect	-1.540e-01	1.625e+00	-0.095	0.924513
## school_pntgF:Shapesqu_rect	1.249e+00	3.191e+00	0.391	0.695635
## school_pntgG:Shapesqu_rect	NA	NA	NA	NA
## school_pntgI:Shapesqu_rect	NA	NA	NA	NA
## school_pntgS:Shapesqu_rect	NA	NA	NA	NA
## school_pntgX:Shapesqu_rect	NA	NA	NA	NA
## school_pntgD/FL:Surface	-9.744e-04	1.398e-03	-0.697	0.485902
## school_pntgF:Surface	-1.055e-03	1.396e-03	-0.756	0.449974
## school_pntgG:Surface	NA	NA	NA	NA
## school_pntgI:Surface	-9.442e-04	1.400e-03	-0.674	0.500184
## school_pntgS:Surface	8.396e-04	2.389e-03	0.351	0.725367
## school_pntgX:Surface	NA	NA	NA	NA
## school_pntgD/FL:engraved1	4.927e-01	6.844e-01	0.720	0.471721
## school_pntgF:engraved1	9.062e-01	6.956e-01	1.303	0.192919
## school_pntgG:engraved1	NA	NA	NA	NA
## school_pntgI:engraved1	NA	NA	NA	NA
## school_pntgS:engraved1	NA	NA	NA	NA
## school_pntgX:engraved1	NA	NA	NA	NA
## school_pntgD/FL:prevcoll1	-5.141e-01	5.500e-01	-0.935	0.350069
## school_pntgF:prevcoll1	-3.277e-01	6.478e-01	-0.506	0.613032
## school_pntgG:prevcoll1	NA	NA	NA	NA
## school_pntgI:prevcoll1	NA	NA	NA	NA
## school_pntgS:prevcoll1	NA	NA	NA	NA
## school_pntgX:prevcoll1	NA	NA	NA	NA
## school_pntgD/FL:paired1	-4.553e-01	8.958e-01	-0.508	0.611385
## school_pntgF:paired1	-6.532e-01	9.057e-01	-0.721	0.470910
## school_pntgG:paired1	NA	NA	NA	NA
## school_pntgI:paired1	-6.810e-03	9.172e-01	-0.007	0.994077
## school_pntgS:paired1	NA	NA	NA	NA
## school_pntgX:paired1	NA	NA	NA	NA
## school_pntgD/FL:finished1	-3.501e-02	3.434e-01	-0.102	0.918804
## school_pntgF:finished1	-1.400e-01	3.979e-01	-0.352	0.724979
## school_pntgG:finished1	NA	NA	NA	NA
## school_pntgI:finished1	NA	NA	NA	NA
## school_pntgS:finished1	NA	NA	NA	NA
## school_pntgX:finished1	NA	NA	NA	NA

## school_pntgD/FL:lrgfont1	-1.293e-01	4.734e-01	-0.273	0.784858
## school_pntgF:lrgfont1	-4.077e-01	5.479e-01	-0.744	0.456999
## school_pntgG:lrgfont1	NA	NA	NA	NA
## school_pntgI:lrgfont1	NA	NA	NA	NA
## school_pntgS:lrgfont1	NA	NA	NA	NA
## school_pntgX:lrgfont1	NA	NA	NA	NA
## school_pntgD/FL:portrait1	-1.033e-01	1.508e+00	-0.069	0.945390
## school_pntgF:portrait1	-1.465e+00	1.422e+00	-1.030	0.303229
## school_pntgG:portrait1	NA	NA	NA	NA
## school_pntgI:portrait1	-3.168e-01	1.630e+00	-0.194	0.845898
## school_pntgS:portrait1	NA	NA	NA	NA
## school_pntgX:portrait1	NA	NA	NA	NA
## school_pntgD/FL:discauth1	5.396e-01	5.443e-01	0.991	0.321693
## school_pntgF:discauth1	2.943e-01	6.077e-01	0.484	0.628284
## school_pntgG:discauth1	NA	NA	NA	NA
## school_pntgI:discauth1	NA	NA	NA	NA
## school_pntgS:discauth1	NA	NA	NA	NA
## school_pntgX:discauth1	NA	NA	NA	NA
## school_pntgD/FL:still_life1	7.680e-01	1.333e+00	0.576	0.564571
## school_pntgF:still_life1	-1.182e-02	1.348e+00	-0.009	0.993007
## school_pntgG:still_life1	NA	NA	NA	NA
## school_pntgI:still_life1	-1.013e-01	1.535e+00	-0.066	0.947412
## school_pntgS:still_life1	NA	NA	NA	NA
## school_pntgX:still_life1	NA	NA	NA	NA
## diff_origin1:artistliving1	5.449e-01	3.798e-01	1.435	0.151665
## diff_origin1:endbuyerC	1.073e+00	1.345e+00	0.798	0.424975
## diff_origin1:endbuyerD	1.568e+00	1.329e+00	1.180	0.238160
## diff_origin1:endbuyerE	2.189e+00	1.361e+00	1.609	0.107825
## diff_origin1:endbuyerU	2.043e+00	1.351e+00	1.512	0.130692
## diff_origin1:endbuyerX	1.786e+00	1.342e+00	1.331	0.183594
## diff_origin1:authorstyle1	4.936e-01	6.000e-01	0.823	0.410805
## diff_origin1:Interm1	5.620e-01	6.507e-01	0.864	0.387892
## diff_origin1:Shapeoval	NA	NA	NA	NA
## diff_origin1:Shaperound	-7.072e-01	1.445e+00	-0.489	0.624760
## diff_origin1:Shapesqu_rect	-1.036e+00	1.141e+00	-0.908	0.364209
## diff_origin1:Surface	-3.026e-04	1.354e-04	-2.234	0.025677 *
## diff_origin1:engraved1	-3.268e-01	5.560e-01	-0.588	0.556781
## diff_origin1:prevcoll1	2.110e-02	7.535e-01	0.028	0.977663
## diff_origin1:paired1	-4.230e-01	2.316e-01	-1.827	0.068009 .
## diff_origin1:finished1	2.223e-01	4.325e-01	0.514	0.607331
## diff_origin1:lrgfont1	-3.645e-01	6.097e-01	-0.598	0.550132
## diff_origin1:portrait1	1.536e+00	7.447e-01	2.063	0.039309 *
## diff_origin1:discauth1	5.989e-01	7.896e-01	0.758	0.448317
## diff_origin1:still_life1	-9.692e-01	7.048e-01	-1.375	0.169330
## artistliving1:endbuyerC	2.140e+00	1.114e+00	1.921	0.054974 .
## artistliving1:endbuyerD	1.771e+00	1.105e+00	1.603	0.109240
## artistliving1:endbuyerE	3.114e+00	1.168e+00	2.667	0.007750 **
## artistliving1:endbuyerU	2.068e+00	1.126e+00	1.836	0.066556 .
## artistliving1:endbuyerX	2.042e+00	1.132e+00	1.803	0.071573 .
## artistliving1:authorstyle1	2.182e+00	1.037e+00	2.104	0.035618 *
## artistliving1:Interm1	-6.395e-01	6.775e-01	-0.944	0.345378
## artistliving1:Shapeoval	NA	NA	NA	NA
## artistliving1:Shaperound	3.033e+00	3.823e+00	0.793	0.427790
## artistliving1:Shapesqu_rect	-1.151e+00	3.096e+00	-0.372	0.710280

## artistliving1:Surface	2.528e-04	1.499e-04	1.687	0.091874	.
## artistliving1:engraved1	-4.261e-01	5.151e-01	-0.827	0.408249	
## artistliving1:prevcoll1	-1.097e+00	8.161e-01	-1.344	0.179209	
## artistliving1:paired1	6.146e-02	2.434e-01	0.253	0.800687	
## artistliving1:finished1	-1.406e-01	3.743e-01	-0.376	0.707143	
## artistliving1:lrgfont1	2.527e-01	5.398e-01	0.468	0.639711	
## artistliving1:portrait1	-1.026e-01	1.223e+00	-0.084	0.933131	
## artistliving1:discauth1	2.203e-01	6.496e-01	0.339	0.734591	
## artistliving1:still_lifel	6.478e-01	5.693e-01	1.138	0.255438	
## endbuyerC:authorstyle1	7.181e-01	7.222e-01	0.994	0.320274	
## endbuyerD:authorstyle1	-2.485e-01	5.332e-01	-0.466	0.641264	
## endbuyerE:authorstyle1	3.886e-01	7.686e-01	0.506	0.613235	
## endbuyerU:authorstyle1	-2.608e-01	6.314e-01	-0.413	0.679612	
## endbuyerX:authorstyle1	NA	NA	NA	NA	
## endbuyerC:Interm1	-6.487e-01	1.426e+00	-0.455	0.649254	
## endbuyerD:Interm1	-1.394e+00	1.543e+00	-0.903	0.366578	
## endbuyerE:Interm1	NA	NA	NA	NA	
## endbuyerU:Interm1	NA	NA	NA	NA	
## endbuyerX:Interm1	NA	NA	NA	NA	
## endbuyerC:Shapeoval	7.224e+00	4.494e+00	1.607	0.108236	
## endbuyerD:Shapeoval	4.538e+00	7.626e+00	0.595	0.551916	
## endbuyerE:Shapeoval	9.261e+00	4.568e+00	2.028	0.042832	*
## endbuyerU:Shapeoval	NA	NA	NA	NA	
## endbuyerX:Shapeoval	NA	NA	NA	NA	
## endbuyerC:Shaperound	3.235e+00	4.155e+00	0.778	0.436456	
## endbuyerD:Shaperound	4.476e+00	4.331e+00	1.033	0.301585	
## endbuyerE:Shaperound	4.935e+00	4.010e+00	1.230	0.218780	
## endbuyerU:Shaperound	1.565e+00	3.819e+00	0.410	0.682049	
## endbuyerX:Shaperound	NA	NA	NA	NA	
## endbuyerC:Shapesqu_rect	3.263e+00	3.910e+00	0.835	0.404075	
## endbuyerD:Shapesqu_rect	3.907e+00	4.115e+00	0.949	0.342615	
## endbuyerE:Shapesqu_rect	4.834e+00	3.777e+00	1.280	0.200782	
## endbuyerU:Shapesqu_rect	2.166e+00	3.526e+00	0.614	0.539205	
## endbuyerX:Shapesqu_rect	NA	NA	NA	NA	
## endbuyerC:Surface	-1.225e-03	1.182e-03	-1.036	0.300557	
## endbuyerD:Surface	-1.090e-03	1.174e-03	-0.928	0.353429	
## endbuyerE:Surface	-1.088e-03	1.191e-03	-0.913	0.361332	
## endbuyerU:Surface	-6.184e-04	1.204e-03	-0.513	0.607753	
## endbuyerX:Surface	-1.095e-03	1.193e-03	-0.918	0.358773	
## endbuyerC:engraved1	1.700e+00	8.508e-01	1.999	0.045873	*
## endbuyerD:engraved1	1.566e+00	7.990e-01	1.960	0.050183	.
## endbuyerE:engraved1	1.822e+00	8.384e-01	2.173	0.029942	*
## endbuyerU:engraved1	1.287e+00	8.549e-01	1.505	0.132523	
## endbuyerX:engraved1	NA	NA	NA	NA	
## endbuyerC:prevcoll1	6.790e-01	9.878e-01	0.687	0.491946	
## endbuyerD:prevcoll1	3.127e-01	9.126e-01	0.343	0.731903	
## endbuyerE:prevcoll1	5.059e-01	9.206e-01	0.550	0.582739	
## endbuyerU:prevcoll1	4.548e-01	1.297e+00	0.351	0.725796	
## endbuyerX:prevcoll1	NA	NA	NA	NA	
## endbuyerC:paired1	-1.650e+00	1.565e+00	-1.054	0.292122	
## endbuyerD:paired1	-1.598e+00	1.560e+00	-1.024	0.305944	
## endbuyerE:paired1	-1.567e+00	1.582e+00	-0.991	0.322055	
## endbuyerU:paired1	-1.443e+00	1.572e+00	-0.918	0.358801	
## endbuyerX:paired1	-1.536e+00	1.572e+00	-0.977	0.328593	

## endbuyerC:finished1	1.632e+00	1.951e+00	0.837	0.403005
## endbuyerD:finished1	1.428e+00	1.948e+00	0.733	0.463747
## endbuyerE:finished1	1.666e+00	1.986e+00	0.839	0.401694
## endbuyerU:finished1	7.766e-01	1.990e+00	0.390	0.696457
## endbuyerX:finished1	5.822e-01	1.996e+00	0.292	0.770566
## endbuyerC:lrgfont1	1.939e+00	1.439e+00	1.348	0.178035
## endbuyerD:lrgfont1	1.685e+00	1.438e+00	1.172	0.241535
## endbuyerE:lrgfont1	1.739e+00	1.624e+00	1.071	0.284506
## endbuyerU:lrgfont1	2.166e+00	1.529e+00	1.417	0.156886
## endbuyerX:lrgfont1	1.984e+00	1.609e+00	1.233	0.217763
## endbuyerC:portrait1	-2.502e-01	1.026e+00	-0.244	0.807493
## endbuyerD:portrait1	-4.196e-01	8.217e-01	-0.511	0.609680
## endbuyerE:portrait1	-2.844e-01	8.219e-01	-0.346	0.729340
## endbuyerU:portrait1	-4.039e-01	1.021e+00	-0.395	0.692611
## endbuyerX:portrait1	NA	NA	NA	NA
## endbuyerC:discauth1	-5.222e-01	8.325e-01	-0.627	0.530601
## endbuyerD:discauth1	-2.284e-01	5.544e-01	-0.412	0.680432
## endbuyerE:discauth1	5.618e-01	6.809e-01	0.825	0.409453
## endbuyerU:discauth1	-6.628e-01	6.587e-01	-1.006	0.314504
## endbuyerX:discauth1	NA	NA	NA	NA
## endbuyerC:still_life1	-6.855e-01	1.028e+00	-0.667	0.504952
## endbuyerD:still_life1	2.431e-01	7.542e-01	0.322	0.747287
## endbuyerE:still_life1	1.728e+00	1.303e+00	1.325	0.185303
## endbuyerU:still_life1	-9.968e-01	7.170e-01	-1.390	0.164674
## endbuyerX:still_life1	NA	NA	NA	NA
## authorstyle1:Interm1	NA	NA	NA	NA
## authorstyle1:Shapeoval	NA	NA	NA	NA
## authorstyle1:Shaperound	-1.585e+00	2.331e+00	-0.680	0.496536
## authorstyle1:Shapesqu_rect	NA	NA	NA	NA
## authorstyle1:Surface	1.655e-04	2.555e-04	0.648	0.517197
## authorstyle1:engraved1	2.303e+00	1.578e+00	1.459	0.144789
## authorstyle1:prevcoll1	-5.399e+00	2.684e+00	-2.011	0.044529 *
## authorstyle1:paired1	-1.251e-01	4.221e-01	-0.296	0.766956
## authorstyle1:finished1	9.884e-01	1.165e+00	0.848	0.396364
## authorstyle1:lrgfont1	NA	NA	NA	NA
## authorstyle1:portrait1	-2.280e+00	1.859e+00	-1.227	0.220106
## authorstyle1:discauth1	-9.918e-01	9.166e-01	-1.082	0.279442
## authorstyle1:still_life1	NA	NA	NA	NA
## Interm1:Shapeoval	-1.720e+00	5.222e+00	-0.329	0.741974
## Interm1:Shaperound	2.728e-01	1.271e+00	0.215	0.830073
## Interm1:Shapesqu_rect	NA	NA	NA	NA
## Interm1:Surface	4.330e-04	2.352e-04	1.841	0.065915 .
## Interm1:engraved1	-3.340e-02	5.836e-01	-0.057	0.954371
## Interm1:prevcoll1	-9.672e-01	5.884e-01	-1.644	0.100462
## Interm1:paired1	-2.347e-01	3.508e-01	-0.669	0.503696
## Interm1:finished1	-6.967e-02	4.216e-01	-0.165	0.868778
## Interm1:lrgfont1	-9.635e-01	3.978e-01	-2.422	0.015574 *
## Interm1:portrait1	-1.133e+00	1.204e+00	-0.941	0.346758
## Interm1:discauth1	1.075e+00	8.008e-01	1.343	0.179687
## Interm1:still_life1	1.571e+00	1.006e+00	1.562	0.118510
## Shapeoval:Surface	-3.930e-03	6.383e-03	-0.616	0.538179
## Shaperound:Surface	-1.413e-03	3.361e-03	-0.420	0.674272
## Shapesqu_rect:Surface	-9.705e-04	3.336e-03	-0.291	0.771144
## Shapeoval:engraved1	-9.931e-01	1.819e+00	-0.546	0.585123

## Shaperound:engraved1	NA	NA	NA	NA
## Shapesqu_rect:engraved1	NA	NA	NA	NA
## Shapeoval:prevcoll1	-3.821e+00	7.464e+00	-0.512	0.608818
## Shaperound:prevcoll1	5.689e-01	1.383e+00	0.411	0.680955
## Shapesqu_rect:prevcoll1	NA	NA	NA	NA
## Shapeoval:paired1	-1.162e+00	1.645e+00	-0.706	0.480148
## Shaperound:paired1	-4.104e-04	1.094e+00	0.000	0.999701
## Shapesqu_rect:paired1	NA	NA	NA	NA
## Shapeoval:finished1	-1.768e+00	2.212e+00	-0.799	0.424340
## Shaperound:finished1	-1.648e+00	3.029e+00	-0.544	0.586515
## Shapesqu_rect:finished1	-1.806e+00	1.736e+00	-1.041	0.298279
## Shapeoval:lrgfont1	NA	NA	NA	NA
## Shaperound:lrgfont1	NA	NA	NA	NA
## Shapesqu_rect:lrgfont1	NA	NA	NA	NA
## Shapeoval:portrait1	-1.595e+00	3.208e+00	-0.497	0.619109
## Shaperound:portrait1	NA	NA	NA	NA
## Shapesqu_rect:portrait1	NA	NA	NA	NA
## Shapeoval:discauth1	NA	NA	NA	NA
## Shaperound:discauth1	NA	NA	NA	NA
## Shapesqu_rect:discauth1	-1.619e+00	1.621e+00	-0.999	0.318193
## Shapeoval:still_life1	-2.288e+00	2.644e+00	-0.865	0.387113
## Shaperound:still_life1	NA	NA	NA	NA
## Shapesqu_rect:still_life1	NA	NA	NA	NA
## Surface:engraved1	-3.036e-04	2.224e-04	-1.365	0.172449
## Surface:prevcoll1	1.546e-05	2.408e-04	0.064	0.948809
## Surface:paired1	4.031e-05	1.007e-04	0.400	0.689105
## Surface:finished1	7.563e-05	1.713e-04	0.441	0.658955
## Surface:lrgfont1	-1.347e-04	2.286e-04	-0.589	0.555903
## Surface:portrait1	8.615e-05	3.712e-04	0.232	0.816532
## Surface:discauth1	-5.320e-04	2.968e-04	-1.792	0.073340
## Surface:still_life1	-7.183e-04	3.498e-04	-2.053	0.040275 *
## engraved1:prevcoll1	2.060e-01	5.989e-01	0.344	0.730907
## engraved1:paired1	4.213e-02	4.051e-01	0.104	0.917187
## engraved1:finished1	-1.345e-01	4.645e-01	-0.290	0.772228
## engraved1:lrgfont1	1.916e-01	5.042e-01	0.380	0.703968
## engraved1:portrait1	1.555e-03	1.665e+00	0.001	0.999255
## engraved1:discauth1	8.389e-01	6.154e-01	1.363	0.173028
## engraved1:still_life1	NA	NA	NA	NA
## prevcoll1:paired1	5.442e-01	4.343e-01	1.253	0.210444
## prevcoll1:finished1	-9.424e-01	4.849e-01	-1.943	0.052198
## prevcoll1:lrgfont1	2.722e-01	5.147e-01	0.529	0.596992
## prevcoll1:portrait1	3.164e-01	9.122e-01	0.347	0.728725
## prevcoll1:discauth1	-7.485e-01	7.819e-01	-0.957	0.338602
## prevcoll1:still_life1	NA	NA	NA	NA
## paired1:finished1	3.231e-01	2.628e-01	1.230	0.219046
## paired1:lrgfont1	-7.758e-01	3.204e-01	-2.422	0.015601 *
## paired1:portrait1	2.268e-01	6.101e-01	0.372	0.710100
## paired1:discauth1	-7.715e-01	4.781e-01	-1.614	0.106891
## paired1:still_life1	-6.608e-01	4.992e-01	-1.324	0.185837
## finished1:lrgfont1	-2.193e-01	3.230e-01	-0.679	0.497239
## finished1:portrait1	-1.208e+00	8.258e-01	-1.462	0.143882
## finished1:discauth1	4.303e-01	3.797e-01	1.133	0.257363
## finished1:still_life1	-1.101e-01	1.065e+00	-0.103	0.917708
## lrgfont1:portrait1	2.894e-02	8.926e-01	0.032	0.974144

```
## lrgfont1:discauth1      -3.781e-01  5.620e-01  -0.673  0.501246
## lrgfont1:still_life1    8.605e-01  1.104e+00   0.780  0.435781
## portrait1:discauth1     1.007e+00  1.056e+00   0.954  0.340423
## portrait1:still_life1      NA         NA         NA         NA
## discauth1:still_life1   -6.526e-02  1.612e+00  -0.040  0.967726
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.188 on 1197 degrees of freedom
## Multiple R-squared:  0.694, Adjusted R-squared:  0.6167
## F-statistic: 8.988 on 302 and 1197 DF, p-value: < 2.2e-16
```

Model Selection

After completing the initial exploratory data analysis, methods including Stepwise Best Subset Selection using both AIC and BIC were used in order to assess more systematically which covariates were most important for predicting the logprice of paintings. While the number of relevant covariates was initially thinned by examining the data and determining which variables were best suited for modeling (e.g. via dimension reduction, elimination or recoding of categorical variables with too many levels or too few observations for a given level to be useful in estimating a coefficient), there still remained a large number of covariates from which to choose. The goal in using the above described methodology was to demonstrate among several methods, both frequentist and Bayesian, which covariates were routinely deemed to be the most important for modeling logprice.

The variable selection methods described above remain computationally intensive, particularly given the number of variables and potential two-way interactions that must be considered. In order to begin the analysis, The two-way interactions were considered using stepwise selection (AIC & BIC). The goal of this penalized selection process was to avoid overfitting and to deliver a model that was both interpretable and performed well in prediction. Then we compared the results from two methods and filtered out interactions that have NAs as coefficients, that are not significant, and that do not make sense to be interacted (such as *artistling * endbuyer*).

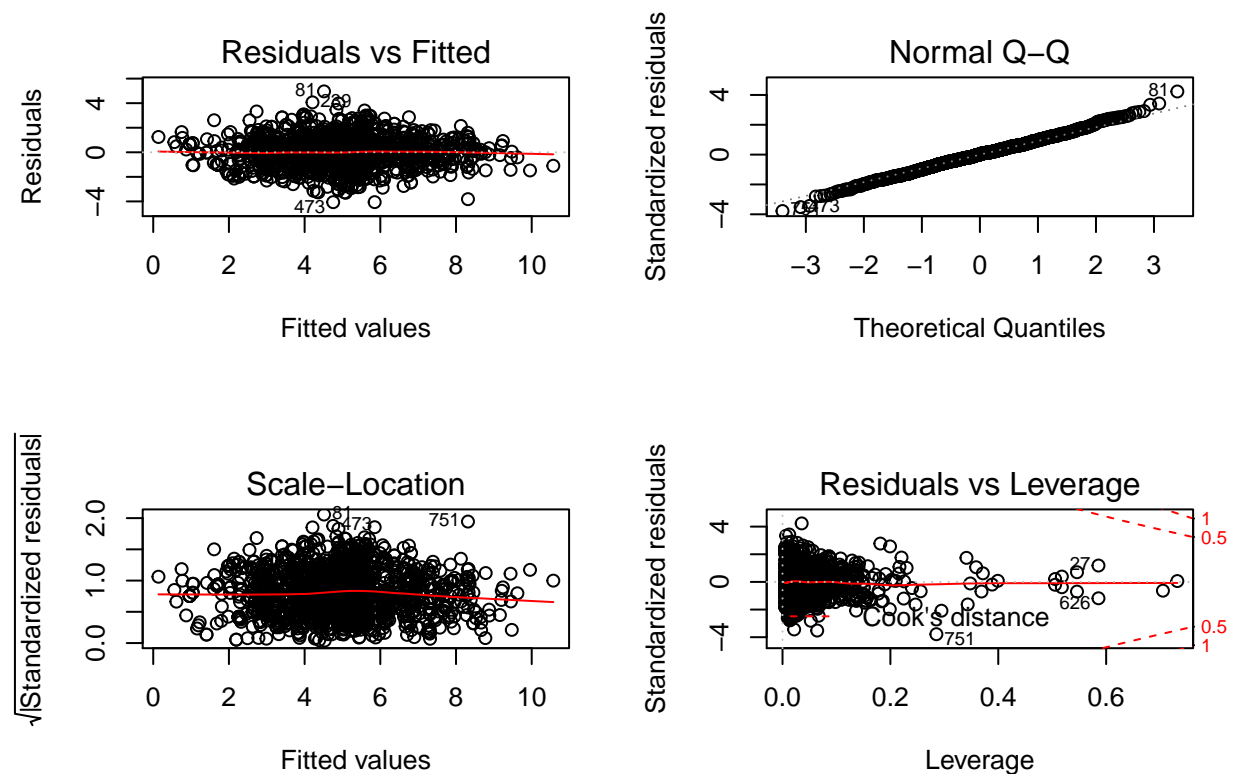
Ultimately, the following variables were selected using the above methods and were fit using OLS regression. The R^2 reduces to 0.6269, which is expected. All the estimated coefficients do not contain NAs.

```
##
## Call:
## lm(formula = logprice ~ Shape + school_pntg + dealer * Interm +
##   dealer * Surface + dealer * paired + dealer * finished +
##   dealer * discauth + diff_origin * Surface + diff_origin *
##   portrait + artistliving * endbuyer + artistliving * authorstyle +
##   Interm * Surface + Interm * lrgfont + Surface * lrgfont +
##   Surface * still_life + Surface * discauth + prevcoll * finished +
##   paired * lrgfont + paired * discauth + diff_origin * authorstyle +
##   diff_origin * still_life + finished * discauth + lrgfont *
##   discauth + artistliving * finished + Interm * portrait, data = paintings_train_2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.0616 -0.7173  0.0137  0.7230  4.9598
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.212e+00  1.315e+00   3.202  0.00139 **
## Shapeoval     5.778e-01  3.765e-01   1.535  0.12512
```

## Shaperound	-5.042e-01	3.445e-01	-1.464	0.14353	
## Shapesqu_rect	5.298e-01	2.544e-01	2.082	0.03751	*
## school_pntgD/FL	-6.702e-01	1.220e+00	-0.549	0.58276	
## school_pntgF	-1.359e+00	1.221e+00	-1.113	0.26579	
## school_pntgG	-3.061e+00	1.717e+00	-1.783	0.07474	.
## school_pntgI	-1.305e+00	1.223e+00	-1.067	0.28632	
## school_pntgS	-8.490e-01	1.515e+00	-0.560	0.57541	
## school_pntgX	-1.861e+00	1.235e+00	-1.507	0.13206	
## dealerL	2.465e+00	2.091e-01	11.791	< 2e-16	***
## dealerP	1.232e+00	2.790e-01	4.417	1.08e-05	***
## dealerR	1.784e+00	1.830e-01	9.751	< 2e-16	***
## Interm1	-1.155e+00	4.790e-01	-2.411	0.01602	*
## Surface	8.369e-04	1.974e-04	4.239	2.39e-05	***
## paired1	2.440e-01	1.896e-01	1.287	0.19834	
## finished1	1.100e+00	2.213e-01	4.971	7.46e-07	***
## discauth1	1.101e+00	3.881e-01	2.836	0.00463	**
## diff_origin1	-5.321e-01	1.061e-01	-5.013	6.01e-07	***
## portrait1	-9.168e-01	2.058e-01	-4.454	9.09e-06	***
## artistliving1	-6.625e-01	7.850e-01	-0.844	0.39889	
## endbuyerC	-3.484e-01	3.748e-01	-0.930	0.35273	
## endbuyerD	-3.455e-01	3.705e-01	-0.933	0.35119	
## endbuyerE	-1.114e+00	3.834e-01	-2.904	0.00374	**
## endbuyerU	-8.528e-01	3.811e-01	-2.238	0.02541	*
## endbuyerX	-1.553e+00	3.782e-01	-4.107	4.24e-05	***
## authorstyle1	-1.786e+00	4.225e-01	-4.226	2.53e-05	***
## lrgfont1	1.780e+00	1.767e-01	10.072	< 2e-16	***
## still_life1	-1.287e-01	2.421e-01	-0.532	0.59507	
## prevcoll1	1.488e+00	1.689e-01	8.808	< 2e-16	***
## dealerL:Interm1	1.516e+00	8.543e-01	1.774	0.07624	.
## dealerP:Interm1	1.711e+00	1.316e+00	1.301	0.19354	
## dealerR:Interm1	2.438e+00	4.934e-01	4.940	8.72e-07	***
## dealerL:Surface	-5.210e-04	2.076e-04	-2.510	0.01219	*
## dealerP:Surface	-3.382e-04	3.425e-04	-0.987	0.32363	
## dealerR:Surface	-5.428e-04	1.992e-04	-2.725	0.00651	**
## dealerL:paired1	-1.259e+00	2.541e-01	-4.956	8.04e-07	***
## dealerP:paired1	-7.507e-01	3.524e-01	-2.130	0.03332	*
## dealerR:paired1	-3.499e-01	2.080e-01	-1.682	0.09272	.
## dealerL:finished1	-4.905e-01	4.979e-01	-0.985	0.32466	
## dealerP:finished1	-6.205e-01	4.074e-01	-1.523	0.12798	
## dealerR:finished1	-5.430e-01	2.422e-01	-2.242	0.02509	*
## dealerL:discauth1	-4.222e-02	9.221e-01	-0.046	0.96349	
## dealerP:discauth1	-1.851e+00	9.470e-01	-1.955	0.05078	.
## dealerR:discauth1	-1.025e+00	3.607e-01	-2.843	0.00454	**
## Surface:diff_origin1	-1.848e-04	9.509e-05	-1.944	0.05212	.
## diff_origin1:portrait1	8.941e-01	4.111e-01	2.175	0.02978	*
## artistliving1:endbuyerC	1.040e+00	8.124e-01	1.281	0.20052	
## artistliving1:endbuyerD	1.174e+00	8.014e-01	1.465	0.14318	
## artistliving1:endbuyerE	2.149e+00	8.575e-01	2.506	0.01232	*
## artistliving1:endbuyerU	1.299e+00	8.249e-01	1.575	0.11553	
## artistliving1:endbuyerX	1.486e+00	8.084e-01	1.838	0.06621	.
## artistliving1:authorstyle1	9.950e-01	8.818e-01	1.128	0.25933	
## Interm1:Surface	4.252e-04	1.575e-04	2.700	0.00701	**
## Interm1:lrgfont1	-8.576e-01	2.704e-01	-3.172	0.00155	**
## Surface:lrgfont1	-2.117e-04	1.153e-04	-1.836	0.06654	.

```
## Surface:still_life1      -5.376e-04  2.489e-04  -2.160  0.03097 *
## Surface:discauth1      -1.755e-04  2.249e-04  -0.780  0.43548
## finished1:prevcoll1    -1.061e+00  3.253e-01  -3.261  0.00114 **
## paired1:lrgfont1       -7.934e-01  2.549e-01  -3.112  0.00189 **
## paired1:discauth1      -3.764e-01  3.423e-01  -1.100  0.27167
## diff_origin1:authorstyle1  9.514e-01  4.518e-01  2.106  0.03540 *
## diff_origin1:still_life1 -8.577e-01  3.490e-01  -2.458  0.01409 *
## finished1:discauth1     5.652e-01  3.110e-01  1.817  0.06941 .
## discauth1:lrgfont1     -5.126e-01  4.392e-01  -1.167  0.24329
## finished1:artistliving1 -3.050e-01  2.802e-01  -1.088  0.27656
## Interml:portrait1      -8.708e-01  5.870e-01  -1.484  0.13815
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.194 on 1433 degrees of freedom
## Multiple R-squared:  0.6299, Adjusted R-squared:  0.6128
## F-statistic: 36.95 on 66 and 1433 DF, p-value: < 2.2e-16
```

Residuals



After fitting the model, we created the four model diagnostic plots. The overall appearances of all four plots seem acceptable, with no obvious outlier or highly influential points shown. The model also does not violate the normality assumption. The constant variance of residuals assumption seems to be satisfied. However, there are 2 cases that are dropped from the plots because they both have leverage of 1, indicating that they could potentially be the outlying cases of underpriced/overpriced paintings that we will later on investigate in, or have extreme price values. It is worth our attention to specifically look at these cases.

Variables

	Coefficient	2.5%	97.5%
(Intercept)	67.516	5.114	891.423
Shapeoval	1.782	0.851	3.730
Shaperound	0.604	0.307	1.187
Shapesqu_rect	1.699	1.031	2.798
school_pntgD/FL	0.512	0.047	5.598
school_pntgF	0.257	0.023	2.817
school_pntgG	0.047	0.002	1.358
school_pntgI	0.271	0.025	2.988
school_pntgS	0.428	0.022	8.362
school_pntgX	0.156	0.014	1.753
dealerL	11.768	7.809	17.735
dealerP	3.428	1.984	5.926
dealerR	5.956	4.160	8.528
Interm1	0.315	0.123	0.806
Surface	1.001	1.000	1.001
paired1	1.276	0.880	1.851
finished1	3.005	1.946	4.638
discauth1	3.006	1.404	6.436
diff_origin1	0.587	0.477	0.723
portrait1	0.400	0.267	0.599
artistliving1	0.516	0.111	2.405
endbuyerC	0.706	0.338	1.472
endbuyerD	0.708	0.342	1.464
endbuyerE	0.328	0.155	0.697
endbuyerU	0.426	0.202	0.900
endbuyerX	0.212	0.101	0.444
authorstyle1	0.168	0.073	0.384
lrgfont1	5.929	4.192	8.386
still_life1	0.879	0.547	1.414
prevcoll1	4.429	3.179	6.169
dealerL:Interm1	4.553	0.852	24.324
dealerP:Interm1	5.537	0.419	73.126
dealerR:Interm1	11.446	4.348	30.130
dealerL:Surface	0.999	0.999	1.000
dealerP:Surface	1.000	0.999	1.000
dealerR:Surface	0.999	0.999	1.000
dealerL:paired1	0.284	0.172	0.467
dealerP:paired1	0.472	0.236	0.942
dealerR:paired1	0.705	0.469	1.060
dealerL:finished1	0.612	0.231	1.626
dealerP:finished1	0.538	0.242	1.196
dealerR:finished1	0.581	0.361	0.934
dealerL:discauth1	0.959	0.157	5.851
dealerP:discauth1	0.157	0.025	1.006
dealerR:discauth1	0.359	0.177	0.728
Surface:diff_origin1	1.000	1.000	1.000
diff_origin1:portrait1	2.445	1.092	5.477
artistliving1:endbuyerC	2.830	0.575	13.930
artistliving1:endbuyerD	3.234	0.672	15.577
artistliving1:endbuyerE	8.576	1.595	46.112

	Coefficient	2.5%	97.5%
artistliving1:endbuyerU	3.666	0.727	18.490
artistliving1:endbuyerX	4.420	0.905	21.583
artistliving1:authorstyle1	2.705	0.480	15.253
Interm1:Surface	1.000	1.000	1.001
Interm1:lrgfont1	0.424	0.250	0.721
Surface:lrgfont1	1.000	1.000	1.000
Surface:still_life1	0.999	0.999	1.000
Surface:discauth1	1.000	0.999	1.000
finished1:prevcoll1	0.346	0.183	0.655
paired1:lrgfont1	0.452	0.274	0.746
paired1:discauth1	0.686	0.351	1.343
diff_origin1:authorstyle1	2.589	1.067	6.282
diff_origin1:still_life1	0.424	0.214	0.841
finished1:discauth1	1.760	0.956	3.239
discauth1:lrgfont1	0.599	0.253	1.417
finished1:artistliving1	0.737	0.425	1.277
Interm1:portrait1	0.419	0.132	1.324

In the linear model we selected, we included Shape, school_pntg, dealer, Interm, Surface, paired, finished, discauth, diff_origin, portrait, artistliving, endbuyer, authorstyle, lrgfont, still_life, and prevcoll as our base predictors. Interactions selected by the model selection process and, for the sake of interpretation, those that are reasonable and interpretable are kept in the model as well. Since the response variable was originally log-transformed, the exponentiated coefficients and confidence intervals are shown in the table.

4. Summary and Conclusions

What is the (median) price for the “baseline” category if there are categorical or dummy variables in the model (add CIs)? (be sure to include units!) Highlight important findings and potential limitations of your model. Does it appear that interactions are important? What are the most important variables and/or interactions? Provide interpretations of how the most important variables influence the (median) price giving a range (CI). Correct interpretation of coefficients for the log model desirable for full points.

Provide recommendations for the art historian about features or combination of features to look for to find the most valuable paintings.