STA 521 Final Project Part II

Team 10: Bin Han, Jingyi Zhang, Jonathan Klus

12 December 2018

Read in Training Data

Data Cleaning

Package Imputation

Package Imputation

1. Introduction

In this study, the auction prices of paintings in 18th century Paris were examined. Specifically, we wish to understand the variables which affect the prices of the paintings, and then be able to predict auction prices based on characteristics of a certain painting. By fitting an appropriate model, we will also be creating a tool to help decide whether specific paintings that are either underpriced or overpriced given their realization of the covariates that were included in the model.

One of the main challenges in building this model is to narrow down the number of covariates from the 59 candidates in the original data set to less than 20 in the final model. This must be done in such a way that an undue amount of bias is not introduced, and overfitting is avoided. Another challenge is to properly deal with the messiness of the data, including both missingness, covariates with a very large number of levels, multicollinearity in the data, and discrepancies in data entries (e.g. same category marked differently).

The ability to explain the results and provide recommendations to individuals without statistical background is equally important and challenging, since the primary audience for this analysis is intended to be art historians. The goal was therefore to balance predictive performance, model simplicity, and interpretability in order to create a pricing model for artwork in 18th century France.

2. Exploratory data analysis

A) Data summary & cleaning

To begin, we looked at the summary of the original training data. There are few numeric variables and a lot of binary variables. Some variables, such as `Interm`, `Surface`, `Height_in` etc. have mising values, which needed to be imputed. The following steps were taken to clean the data:

a. The first step was to reduce the dimensionality of the problem by removing variables, including: `lot`, `sale`, `price`, `count`, `subject`, `authorstandard`, `author`, `winningbidder`, and `other`. From the summary table, the `count` variable has all 1's; the `other` variable does not convey useful information; the other variables, such as `names` and `subjects`, are not useful in predicting the response variable (such as names). From the table of unique values we can see that some variables have thousands of unique values. Therefore, we can remove them in the first step.
b. By further screening the variables, we found out that `Surface` and `Surface_Rnd`, `Surface_Rect` are correlated, 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.

e. The variable `subject` also contains a lot of levels, potentially with many observations representing the same thing but simply expressed in different ways (i.e. different spellings, letter capitalization). Strings of the same meaning are detected from subjects and recatogorized together. Those levels with more than 20 observations are kept while all others are put together under other.

f. On top of the previous steps, we also recategorized the authors since many authors have too few paintings but were still counted as a level in the categorical variable. Only the authors with more than 10 paintings are kept as a distinct level, and all others are merged into the `other` level.

```
kable(paintings_train_2 %>%
  group_by(author) %>%
  summarise(sum = n()) %>%
  arrange(desc(sum)), align = "c")
```

author	sum
other	1307
David Teniers	46
Philippe Wouvermans	27
Francois Boucher	26
Charles de la Fosse	17
French	16
Gasparo Van Vitelle	13
Rosalba Carriera	13
Gaspard Netscher	12
Nicolas Poussin	12
Nicolas Berghem	11

```
kable(paintings_train_2 %>%
  group_by(subject) %>%
  summarise(sum = n()) %>%
  arrange(desc(sum)), align = "c")
```

subject	sum
other	605
Paysage	324
People	194
Saint	121
Portrait	43
Fruit$Flower	42
Adoration	39
Arch	31
Buste	30
Marine	25
Battle	23

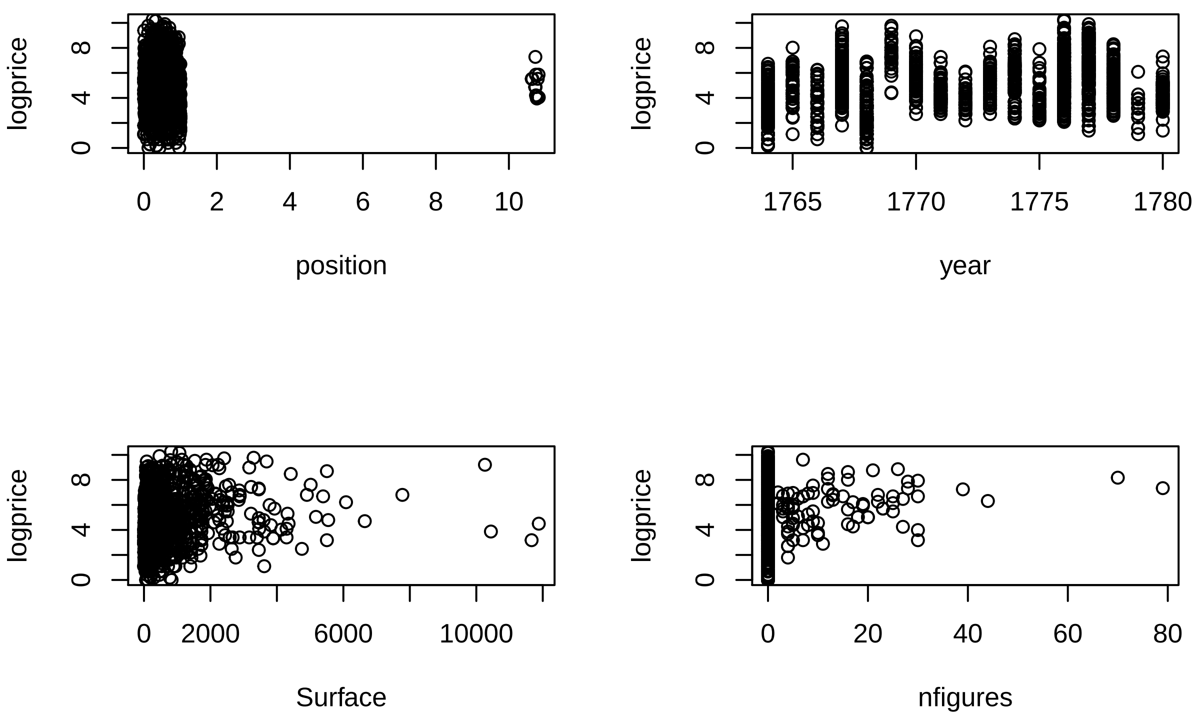
subject	sum
Sujet	23

B). Plots

Then we analyed 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.

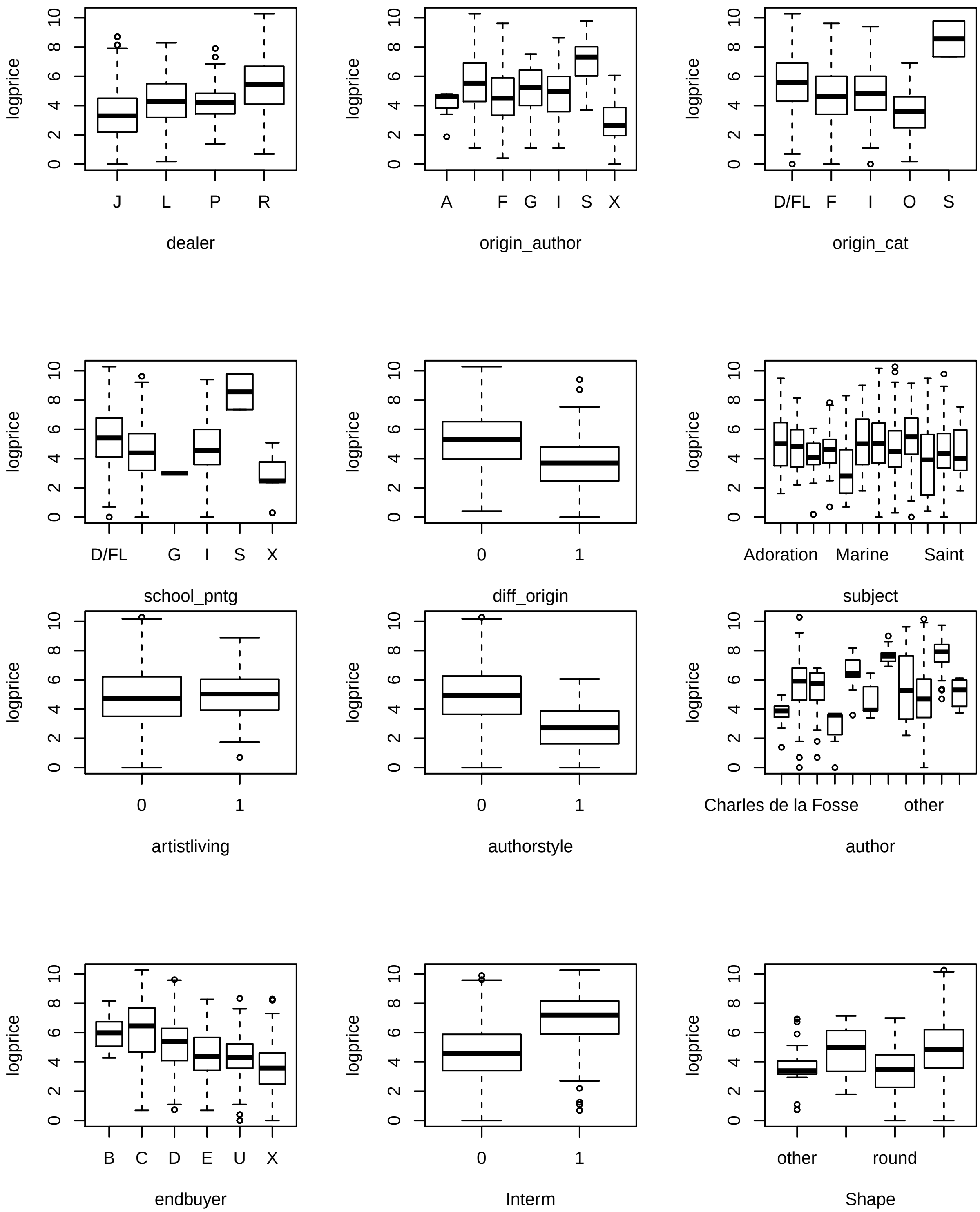
```
## numeric
par(mfrow = c(2, 2))
for (i in 1:ncol(graph_numeric)){
  plot(y = paintings_train_2$logprice,
       x = graph_numeric[,i],
       ylab = "logprice",
       xlab = names(graph_numeric)[i])
}
```



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): `subject`, `author`, `dealer`, `origin_author`, `origin_cat`, `school_pntg`, `diff_origin`, `authorstyle`, `endbuyer`, `Interm`, `Shape`, `materialCat`, `engraved`, `prevcoll`, `figures`, `finished`, `Irgfont`, `othgenre`, `discauth`, and `still_life`.

```
## categorical
par(mfrow = c(2,3))
```

```r
for (i in 1:12){
  boxplot(paintings_train_2$logprice ~ graph_categorical[,i],
          ylab = "logprice",
          xlab = names(graph_categorical)[i])
}
```



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

For numeric variables, we note that `Surface` and `nfigures` appear to have a weak but positive relationship with `logprice`. Since there are several extremely large values in `position` (potential outliers), it is difficult to know if there is a truly useful relationship here between the majority of points and `logprice`. But we will keep it in the initial model for now.

##3. Discussion of preliminary model Part I (5 points)

The overall characteristics of the model that we built in part I were: relatively lower bias, reasonable coverage and higher rmse, comparing to other teams. The methodologies used to conclude at the first model included EDA analyses, BAS, stepwise selection with AIC and BIC, which doesn't do an exhausive search for all possible models, thus the true model and the best model for prediction might not have been captured.

Since there's inherently a trade off between bias and rmse, it is reasonale that we were able to get a bias value on the lower side while rmse unfortunately was on the higher end. There is room for both values, however, to be improved with a better model, potentially a model other than a linear one, or through deeper data cleaning. Next, besides more data cleaning and re-coding, we will be focusing on models like tree/forest methods, as well as more development on linear models.

4. Development of the final model

Final model

```r
summary(ols.2)
```

```
##
## Call:
## lm(formula = logprice ~ Shape + school_pntg + dealer * Interm +
##     dealer * paired + dealer * artistliving + dealer * diff_origin +
##     artistliving * endbuyer + artistliving * finished + artistliving *
##     year + diff_origin * Surface + diff_origin * portrait + diff_origin *
##     still_life + diff_origin * prevcoll + endbuyer * Surface +
##     endbuyer * paired + endbuyer * year + authorstyle * portrait +
##     Interm * lrgfont + paired * lrgfont + paired * subject +
##     paired * discauth + paired * author + finished * discauth +
##     lrgfont * discauth + prevcoll * dealer + prevcoll * school_pntg +
##     prevcoll * Interm, data = paintings_train_2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.0137 -0.6772  0.0079  0.6553  4.3087
##
## Coefficients: (3 not defined because of singularities)
##                                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      -2.491e+02  1.430e+02  -1.742 0.081815
## Shapeoval                         4.510e-01  3.779e-01   1.193 0.232977
## Shaperound                       -7.104e-01  3.503e-01  -2.028 0.042790
## Shapesqu_rect                     2.498e-01  2.675e-01   0.934 0.350549
## school_pntgF                     -5.571e-01  8.341e-02  -6.679 3.47e-11
## school_pntgG                     -2.384e+00  1.141e+00  -2.089 0.036867
## school_pntgI                     -5.761e-01  1.117e-01  -5.157 2.87e-07
## school_pntgS                      4.817e-01  8.248e-01   0.584 0.559328
## school_pntgX                     -9.379e-01  2.136e-01  -4.391 1.21e-05
```

```
## dealerL                        1.759e+00  1.990e-01   8.837  < 2e-16
## dealerP                        5.264e-01  2.301e-01   2.288 0.022301
## dealerR                        1.828e+00  1.531e-01  11.940  < 2e-16
## Interm1                       -3.867e-01  3.959e-01  -0.977 0.328838
## paired1                        1.149e+00  1.052e+00   1.092 0.274974
## artistliving1                  9.560e+01  4.005e+01   2.387 0.017117
## diff_origin1                   1.703e-01  2.325e-01   0.732 0.464020
## endbuyerC                      5.537e+00  1.447e+02   0.038 0.969484
## endbuyerD                      1.125e+02  1.443e+02   0.780 0.435544
## endbuyerE                     -7.717e+01  1.466e+02  -0.526 0.598714
## endbuyerU                      6.730e+01  1.461e+02   0.461 0.645084
## endbuyerX                      1.008e+02  1.441e+02   0.700 0.484292
## finished1                      6.971e-01  1.012e-01   6.888 8.57e-12
## year                           1.424e-01  8.081e-02   1.763 0.078190
## Surface                        9.423e-04  7.868e-04   1.198 0.231269
## portrait1                     -4.317e-01  2.882e-01  -1.498 0.134378
## still_life1                   -2.981e-01  2.277e-01  -1.309 0.190655
## prevcoll1                      1.338e+00  4.768e-01   2.805 0.005096
## authorstyle1                  -8.918e-01  1.556e-01  -5.730 1.23e-08
## lrgfont1                       1.252e+00  1.560e-01   8.026 2.14e-15
## subjectArch                   -5.646e-01  4.214e-01  -1.340 0.180548
## subjectBattle                 -3.689e-02  4.576e-01  -0.081 0.935769
## subjectBuste                  -6.382e-01  3.349e-01  -1.906 0.056898
## subjectFruit$Flower           -1.587e-01  3.740e-01  -0.424 0.671411
## subjectMarine                 -1.125e-02  3.713e-01  -0.030 0.975829
## subjectother                  -2.865e-02  2.305e-01  -0.124 0.901117
## subjectPaysage                -3.327e-01  2.427e-01  -1.371 0.170660
## subjectPeople                  1.921e-01  2.467e-01   0.779 0.436239
## subjectPortrait               -3.003e-01  4.007e-01  -0.749 0.453799
## subjectSaint                  -1.881e-01  2.521e-01  -0.746 0.455909
## subjectSujet                   1.523e-01  4.005e-01   0.380 0.703718
## discauth1                      3.718e-01  2.011e-01   1.848 0.064747
## authorDavid Teniers            7.696e-01  3.911e-01   1.968 0.049260
## authorFrancois Boucher         3.700e-01  4.166e-01   0.888 0.374652
## authorFrench                  -3.533e-01  9.116e-01  -0.388 0.698423
## authorGaspard Netscher         8.727e-01  4.909e-01   1.778 0.075659
## authorGasparo Van Vitelle      1.279e+00  7.463e-01   1.713 0.086898
## authorNicolas Berghem          1.785e+00  6.080e-01   2.936 0.003375
## authorNicolas Poussin          1.539e+00  4.706e-01   3.270 0.001102
## authorother                    4.947e-01  3.274e-01   1.511 0.131022
## authorPhilippe Wouvermans      1.329e+00  4.431e-01   3.000 0.002751
## authorRosalba Carriera         5.380e-01  6.194e-01   0.869 0.385222
## dealerL:Interm1                1.242e+00  5.533e-01   2.246 0.024891
## dealerP:Interm1                1.125e+00  1.260e+00   0.894 0.371730
## dealerR:Interm1                1.256e+00  4.124e-01   3.047 0.002356
## dealerL:paired1               -7.720e-01  2.635e-01  -2.930 0.003445
## dealerP:paired1               -4.145e-01  3.326e-01  -1.246 0.212865
## dealerR:paired1               -1.066e-01  2.078e-01  -0.513 0.607878
## dealerL:artistliving1         -7.328e-01  3.467e-01  -2.114 0.034728
## dealerP:artistliving1         -6.585e-01  4.257e-01  -1.547 0.122099
## dealerR:artistliving1         -7.038e-01  3.205e-01  -2.196 0.028257
## dealerL:diff_origin1           9.065e-02  2.882e-01   0.315 0.753150
## dealerP:diff_origin1          -5.419e-01  3.466e-01  -1.563 0.118234
## dealerR:diff_origin1          -6.727e-01  2.447e-01  -2.749 0.006058
```

```
## artistliving1:endbuyerC           1.436e+00  8.896e-01   1.614 0.106767
## artistliving1:endbuyerD           1.672e+00  8.806e-01   1.898 0.057865
## artistliving1:endbuyerE           2.042e+00  9.385e-01   2.175 0.029775
## artistliving1:endbuyerU           1.689e+00  9.038e-01   1.869 0.061853
## artistliving1:endbuyerX           1.555e+00  8.990e-01   1.730 0.083884
## artistliving1:finished1          -3.065e-01  2.728e-01  -1.124 0.261318
## artistliving1:year               -5.419e-02  2.252e-02  -2.406 0.016243
## diff_origin1:Surface             -1.281e-04  8.944e-05  -1.433 0.152212
## diff_origin1:portrait1            5.098e-01  4.405e-01   1.157 0.247370
## diff_origin1:still_life1         -1.142e+00  3.430e-01  -3.329 0.000895
## diff_origin1:prevcoll1           -1.691e-02  4.586e-01  -0.037 0.970585
## endbuyerC:Surface                -6.698e-04  7.885e-04  -0.850 0.395734
## endbuyerD:Surface                -6.272e-04  7.886e-04  -0.795 0.426521
## endbuyerE:Surface                -4.286e-04  8.004e-04  -0.535 0.592398
## endbuyerU:Surface                -4.884e-04  8.068e-04  -0.605 0.545028
## endbuyerX:Surface                -6.699e-04  7.896e-04  -0.848 0.396354
## paired1:endbuyerC                -1.361e+00  7.195e-01  -1.892 0.058758
## paired1:endbuyerD                -1.014e+00  7.154e-01  -1.417 0.156655
## paired1:endbuyerE                -6.271e-01  7.502e-01  -0.836 0.403367
## paired1:endbuyerU                -8.798e-01  7.361e-01  -1.195 0.232184
## paired1:endbuyerX                -1.025e+00  7.317e-01  -1.400 0.161654
## endbuyerC:year                   -3.144e-03  8.175e-02  -0.038 0.969325
## endbuyerD:year                   -6.370e-02  8.150e-02  -0.782 0.434610
## endbuyerE:year                    4.309e-02  8.282e-02   0.520 0.602905
## endbuyerU:year                   -3.846e-02  8.252e-02  -0.466 0.641273
## endbuyerX:year                   -5.760e-02  8.137e-02  -0.708 0.479183
## portrait1:authorstyle1            1.372e-01  9.538e-01   0.144 0.885658
## Interm1:lrgfont1                 -2.496e-01  2.464e-01  -1.013 0.311309
## paired1:lrgfont1                 -6.105e-01  2.471e-01  -2.471 0.013594
## paired1:subjectArch               4.646e-01  5.997e-01   0.775 0.438587
## paired1:subjectBattle            -7.584e-01  6.474e-01  -1.171 0.241608
## paired1:subjectBuste              1.482e+00  6.332e-01   2.341 0.019397
## paired1:subjectFruit$Flower       1.544e-01  5.666e-01   0.272 0.785333
## paired1:subjectMarine            -7.256e-01  6.228e-01  -1.165 0.244175
## paired1:subjectother             -1.844e-01  4.173e-01  -0.442 0.658643
## paired1:subjectPaysage            9.678e-02  4.257e-01   0.227 0.820174
## paired1:subjectPeople            -1.681e-01  4.471e-01  -0.376 0.706991
## paired1:subjectPortrait           2.089e-02  5.489e-01   0.038 0.969644
## paired1:subjectSaint              3.356e-01  4.814e-01   0.697 0.485813
## paired1:subjectSujet             -3.540e-01  6.279e-01  -0.564 0.572934
## paired1:discauth1                -3.454e-01  3.214e-01  -1.075 0.282759
## paired1:authorDavid Teniers      -9.459e-01  7.453e-01  -1.269 0.204575
## paired1:authorFrancois Boucher    9.617e-01  8.327e-01   1.155 0.248282
## paired1:authorFrench              5.510e-01  1.138e+00   0.484 0.628295
## paired1:authorGaspard Netscher    5.179e-01  1.110e+00   0.467 0.640772
## paired1:authorGasparo Van Vitelle -5.348e-01  9.964e-01  -0.537 0.591518
## paired1:authorNicolas Berghem     1.166e-01  9.629e-01   0.121 0.903657
## paired1:authorNicolas Poussin     9.131e-02  1.401e+00   0.065 0.948030
## paired1:authorother              -8.722e-02  6.613e-01  -0.132 0.895101
## paired1:authorPhilippe Wouvermans  7.653e-01  8.030e-01   0.953 0.340724
## paired1:authorRosalba Carriera    1.472e-02  9.352e-01   0.016 0.987449
## finished1:discauth1               7.695e-01  2.698e-01   2.852 0.004411
## lrgfont1:discauth1               -7.878e-01  3.524e-01  -2.236 0.025524
## dealerL:prevcoll1                -6.275e-01  6.476e-01  -0.969 0.332744
```

```
## dealerP:prevcoll1              -1.637e+00  8.247e-01  -1.985 0.047367
## dealerR:prevcoll1              -6.603e-01  4.785e-01  -1.380 0.167799
## school_pntgF:prevcoll1          5.039e-01  3.443e-01   1.463 0.143611
## school_pntgG:prevcoll1                 NA         NA      NA       NA
## school_pntgI:prevcoll1          5.474e-01  3.734e-01   1.466 0.142831
## school_pntgS:prevcoll1                 NA         NA      NA       NA
## school_pntgX:prevcoll1                 NA         NA      NA       NA
## Interm1:prevcoll1              -3.125e-01  3.166e-01  -0.987 0.323901
##
## (Intercept)                    .
## Shapeoval
## Shaperound                     *
## Shapesqu_rect
## school_pntgF                   ***
## school_pntgG                   *
## school_pntgI                   ***
## school_pntgS
## school_pntgX                   ***
## dealerL                        ***
## dealerP                        *
## dealerR                        ***
## Interm1
## paired1
## artistliving1                  *
## diff_origin1
## endbuyerC
## endbuyerD
## endbuyerE
## endbuyerU
## endbuyerX
## finished1                      ***
## year                           .
## Surface
## portrait1
## still_life1
## prevcoll1                      **
## authorstyle1                   ***
## lrgfont1                       ***
## subjectArch
## subjectBattle
## subjectBuste                   .
## subjectFruit$Flower
## subjectMarine
## subjectother
## subjectPaysage
## subjectPeople
## subjectPortrait
## subjectSaint
## subjectSujet
## discauth1                      .
## authorDavid Teniers            *
## authorFrancois Boucher
## authorFrench
## authorGaspard Netscher         .
```

```
## authorGasparo Van Vitelle         .
## authorNicolas Berghem             **
## authorNicolas Poussin             **
## authorother
## authorPhilippe Wouvermans         **
## authorRosalba Carriera
## dealerL:Interm1                   *
## dealerP:Interm1
## dealerR:Interm1                   **
## dealerL:paired1                   **
## dealerP:paired1
## dealerR:paired1
## dealerL:artistliving1             *
## dealerP:artistliving1
## dealerR:artistliving1             *
## dealerL:diff_origin1
## dealerP:diff_origin1
## dealerR:diff_origin1              **
## artistliving1:endbuyerC
## artistliving1:endbuyerD           .
## artistliving1:endbuyerE           *
## artistliving1:endbuyerU           .
## artistliving1:endbuyerX           .
## artistliving1:finished1
## artistliving1:year                *
## diff_origin1:Surface
## diff_origin1:portrait1
## diff_origin1:still_life1          ***
## diff_origin1:prevcoll1
## endbuyerC:Surface
## endbuyerD:Surface
## endbuyerE:Surface
## endbuyerU:Surface
## endbuyerX:Surface
## paired1:endbuyerC                 .
## paired1:endbuyerD
## paired1:endbuyerE
## paired1:endbuyerU
## paired1:endbuyerX
## endbuyerC:year
## endbuyerD:year
## endbuyerE:year
## endbuyerU:year
## endbuyerX:year
## portrait1:authorstyle1
## Interm1:lrgfont1
## paired1:lrgfont1                  *
## paired1:subjectArch
## paired1:subjectBattle
## paired1:subjectBuste              *
## paired1:subjectFruit$Flower
## paired1:subjectMarine
## paired1:subjectother
## paired1:subjectPaysage
```

```
## paired1:subjectPeople
## paired1:subjectPortrait
## paired1:subjectSaint
## paired1:subjectSujet
## paired1:discauth1
## paired1:authorDavid Teniers
## paired1:authorFrancois Boucher
## paired1:authorFrench
## paired1:authorGaspard Netscher
## paired1:authorGasparo Van Vitelle
## paired1:authorNicolas Berghem
## paired1:authorNicolas Poussin
## paired1:authorother
## paired1:authorPhilippe Wouvermans
## paired1:authorRosalba Carriera
## finished1:discauth1              **
## lrgfont1:discauth1               *
## dealerL:prevcoll1
## dealerP:prevcoll1                *
## dealerR:prevcoll1
## school_pntgF:prevcoll1
## school_pntgG:prevcoll1
## school_pntgI:prevcoll1
## school_pntgS:prevcoll1
## school_pntgX:prevcoll1
## Interm1:prevcoll1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.113 on 1378 degrees of freedom
## Multiple R-squared:  0.6904, Adjusted R-squared:  0.6632
## F-statistic: 25.39 on 121 and 1378 DF,  p-value: < 2.2e-16
```

Variables: must include an explanation

a. The base variables we chose are: `Shape`, `school_pntg`, `dealer`, `Interm`, `paired`, `artistliving`, `diff_origin`, `endbuyer`, `finished`, `year`, `Surface`, `portrait`, `still_life`, `prevcoll`, `authorstyle`, `lrgfont`, `discauth`, `subject` and `author`.

b. The interactions we used include: `dealer*Interm`, `dealer*paired`, `dealer*artistliving`, `dealer*diff_origin`, `artistliving*endbuyer`, `artistliving*finished`, `artistliving*year`, `diff_origin*Surface`, `diff_origin*portrait`, `diff_origin*still_life`, `diff_origin*prevcoll` `endbuyer*Surface`, `endbuyer*paired`, `endbuyer*year`, `authorstyle*portrait`, `Interm*lrgfont`, `paired*lrgfont`, `paired*subject`, `paired*discauth`, `paired*author`, `finished*discauth`, `lrgfont*discauth`, `prevcoll*dealer`, `prevcoll*school_pntg`, and `prevcoll*Interm`.

c. Partial Explnations:

- dealer: the type of dealer that the auction went through significantly affects the price of the painting. For example, compared with dealer J, the average price from dealer L is `179% higher`. (Same interpretation for dealer P and R, with different coefficients)
- finished: if the painting is noted for being highly finished, the selling price on average is `69.76% higher` than when the painting is not noted for being highly finished.

- prevcoll: when the previous owner is mentioned, the average selling price is `128.0% higher` than when the previous owner is not mentioned.
- lrgfont: when the dealer devotes an additional paragraph, the average selling price is `124.9% higher` than when there is no additional paragraph.
- authorstyle: when the author's name is introduced, the average selling price is expected to be `88.39% lower` than when the author's name is not introduced.
- author: which author painted the painting also has some influence on the price. Compared with author Charles de la Fosse, author David Teniers' paintings are `80.57% higher` in price on average. Author Nicolas Berghem's paintings are `181.4% higher` in price on average.
- dealer&Interm interaction: when an intermediary is present, which the price of the auctioned paintings differs significantly among different dealers. For instance, if the dealer is R and an intermediary is used, the average selling price is `107.0% higher` than when the dealer is J with an intermediary.
- finished*discauth: given that the painting is noted for being highly finished, when the dealer engages with authenticity, the average price is expected to be `76.2% higher`.
- diff_origin:still_life: given that the origin of painting based on nationality of artist is different from the origin of painting based on dealer's classification, if the description indicates still life elements, the price is expected to be `-112.8% lower`.

Variable selection/shrinkage:

a. Linear Model

The linear model from part I does a fairly good job in predicting. Therefore, after adding two more variables in the dataset, we decided to refine the linear model first. The plan was to add new features and interactions into the model, hoping to potentially explain more variation in the response variable. Similar as the process in part I, we applied BMA (Bayesian Model Averaging) to select the base variables that have high posterior probabilities and include them in the initial model. Then we tested all possible interactions and used AIC to select interactions that are good for predicting. However, the output from AIC contain too many interactions, which might lead to the problem of overfitting. Additionally, it contains some interactions with coefficients as NA, and some that do not make sense at all. Therefore I manually removed them and kept twisting around the rest features, which led to the best final linear model in terms of the lowest RMSE.

b. Tree Model

Since we have many categorical variables, as in nature, a tree-based model would be appropriate in addressing the interactions to explain the response variable. We tested two tree models: random forest and boosting (bagging does not work in our case as we have 39 variables, which is beyond the limit of possible selection candidates at each node). As for random forest, we used $mtry = 13$ as this is a regression problem. For boosting, we used 5000 trees and tried different interaction depth (4, 6, 8, 10). Both methods do better job than the linear model from part I, but not as good as the linear model generated above.

c. Poisson&Negative Binomial Regression

The nature of auction price is integer. Even though not strictly count data, but could be treated in such way so that we can use poisson&negative binomial regression. We trained the tested the poisson model first. With all the vairiables included (even with all the interactionss), the residual deviance is 100 times higher than the residual degrees of freedom. We concluded that the poisson model s not appropriate and proceeded with negative binomial model. The RMSE is not better than the linear model is part I (even with all the interactions) and there is still the problem of over-dispersion.

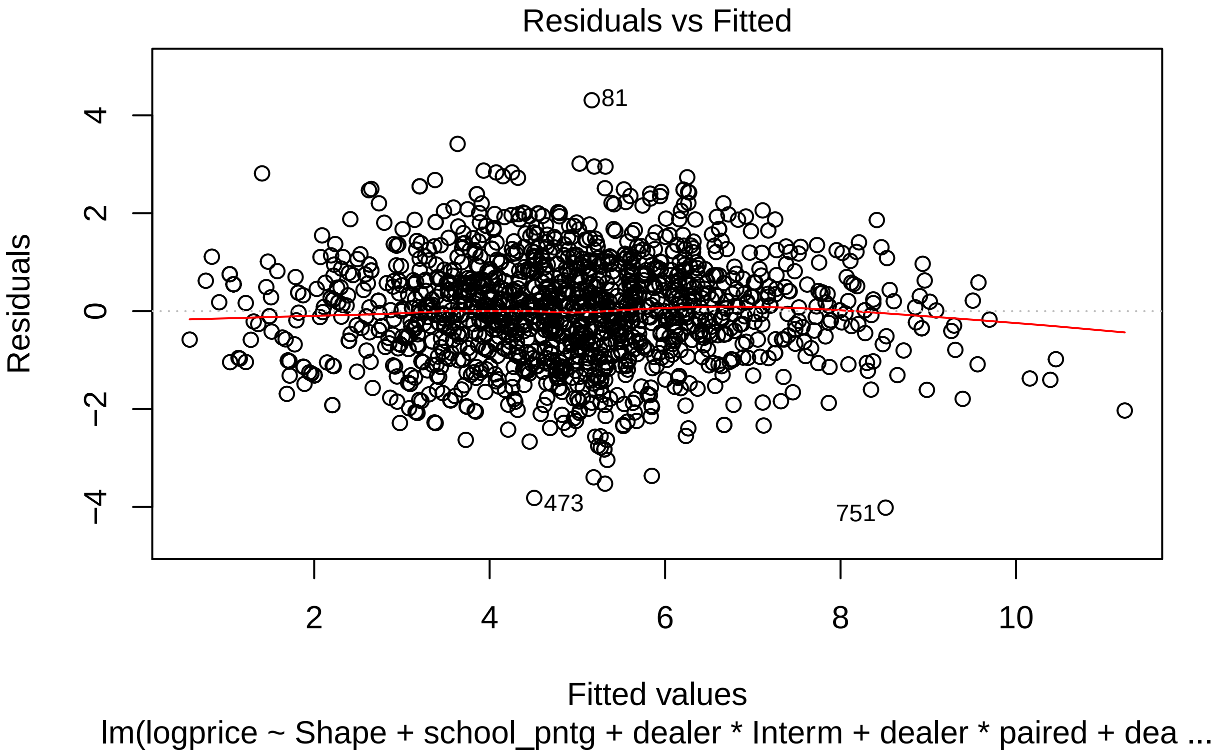
d. Xgboost ## Added some description here

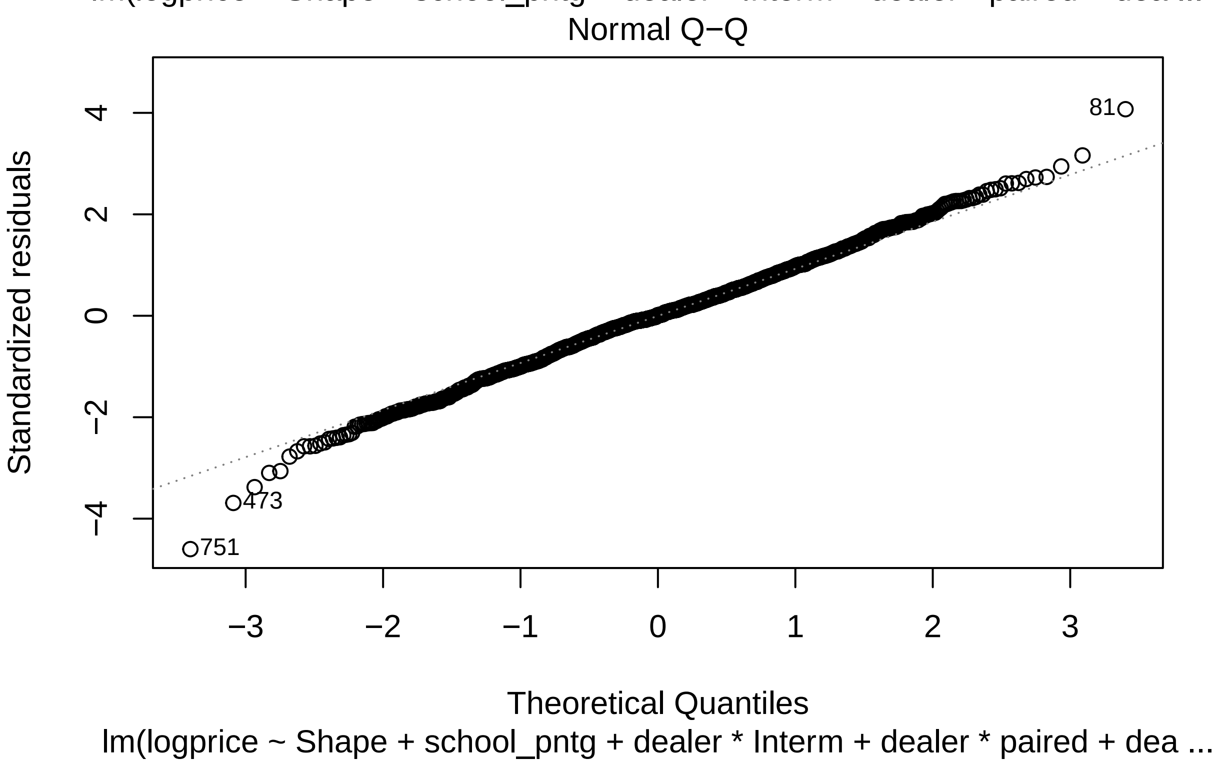
Comparing all the models we fitted above, we conclude that the linear model has the best performance in terms of predicting (lowest rmse). The linear model was relatively more complex than the one in part

I, resuting in some loss in predictability. However, it is still more interpretable than tree-based models. Therefore, we concluded that the best model is the linear model.

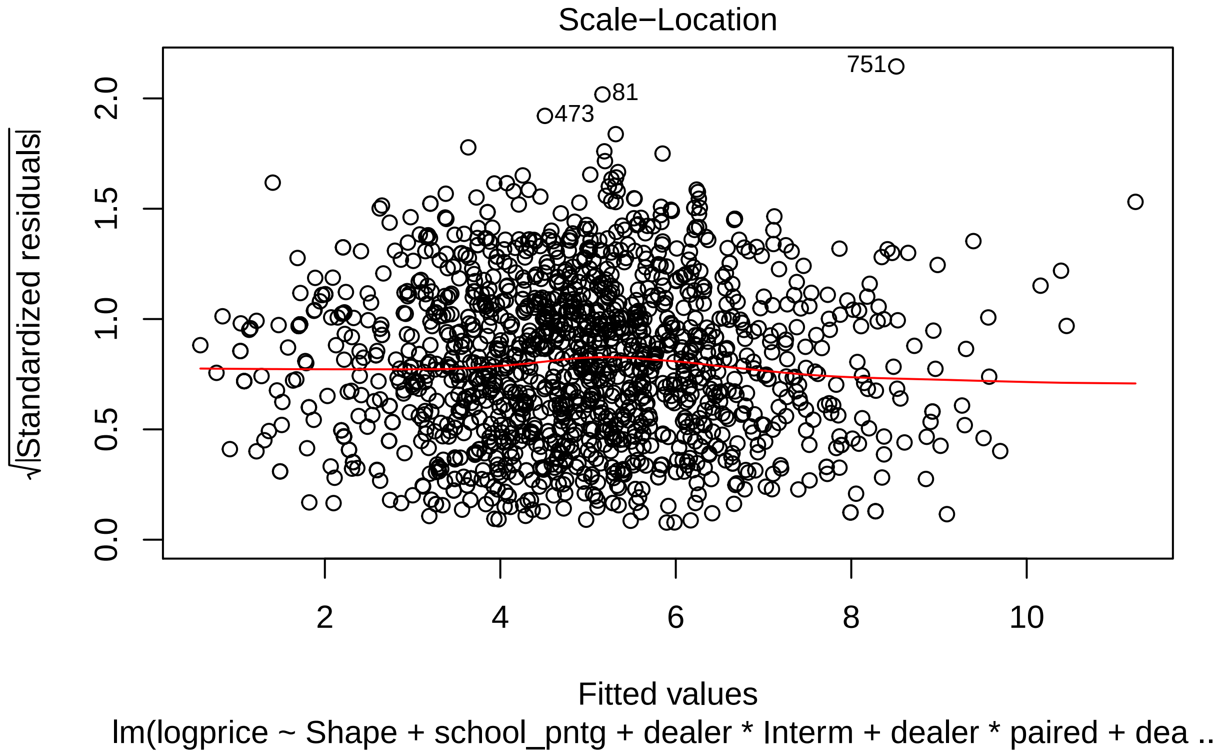
```r
plot(ols.2)
```

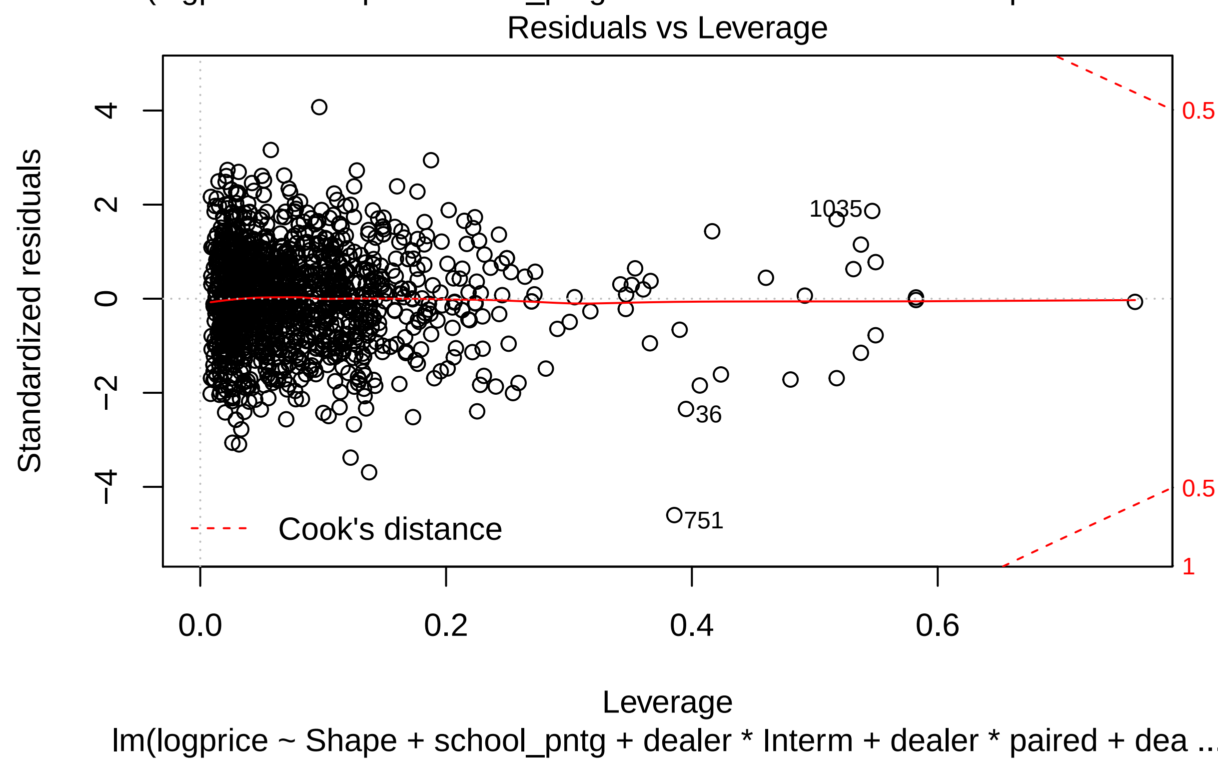
```
## Warning: not plotting observations with leverage one:
##   423, 1129, 1351
```





```
## Warning: not plotting observations with leverage one:
##   423, 1129, 1351
```





Residual: must include a residual plot and a discussion

The residual plot looks fairly good. There seems to have a pattern that the variance is slightly higher with fitted value around 5 and a little lower at the two tails. But in general, the constant vairance assumption is satisfied. The normality assumption is well satisfied from normal qq-plot, with several observations slightly scattered away on the two sides.

discussion of how prediction intervals obtained

```r
PI = data.frame(exp(predict(ols.2,
                 newdata=paintings_test_2,
                 interval = "pred")))
```

```
## Warning in predict.lm(ols.2, newdata = paintings_test_2, interval =
## "pred"): prediction from a rank-deficient fit may be misleading
```

```r
kable(PI, digits = 3, align = "c", caption = "Prediction Interval")
```

Table 3: Prediction Interval

fit	lwr	upr
101.620	11.184	923.346
267.036	19.362	3682.870
171.012	18.031	1621.971
4833.121	418.719	55786.938
461.820	49.260	4329.638
290.560	28.234	2990.147
136.077	14.995	1234.850
4.990	0.497	50.068
113.963	12.135	1070.279
119.101	12.134	1169.063
30.493	3.205	290.086
442.602	48.303	4055.607
426.103	45.231	4014.191
998.617	108.522	9189.206
17.096	1.867	156.574
92.493	10.226	836.616
2181.202	237.890	19999.311
1047.484	113.977	9626.678
105.340	9.611	1154.599
16.426	1.653	163.177
216.039	22.770	2049.724
421.032	31.625	5605.359
15.517	1.521	158.344
32.561	3.508	302.247
12.289	1.303	115.876
6.529	0.684	62.321
173.634	18.316	1646.050
546.504	53.655	5566.442
106.147	11.644	967.656
65.528	7.179	598.104
59.902	6.465	555.017
4708.119	476.563	46513.006

fit	lwr	upr
248.047	25.129	2448.470
423.948	46.630	3854.404
1303.750	131.400	12935.775
830.149	86.266	7988.638
100.559	9.400	1075.734
62.150	6.757	571.654
122.714	13.017	1156.824
50.667	5.390	476.285
119.162	8.360	1698.537
41.931	4.583	383.602
54.458	5.926	500.426
88.656	9.331	842.349
1150.655	113.113	11705.128
2293.337	227.134	23155.495
61.763	6.786	562.178
95.690	10.297	889.238
1636.749	163.362	16398.864
390.094	43.439	3503.110
401.920	44.460	3633.363
39.094	4.060	376.444
380.568	42.035	3445.524
356.095	23.742	5340.887
658.412	69.881	6203.500
612.813	66.690	5631.107
53.103	5.825	484.118
39.800	4.164	380.423
371.971	41.064	3369.407
86.561	9.536	785.753
28.547	2.655	306.961
239.560	26.450	2169.717
258.322	28.566	2336.045
244.279	25.284	2360.046
7805.492	773.133	78803.635
3409.021	334.559	34736.543
50.889	5.128	505.029
94.079	9.950	889.514
359.093	33.835	3811.090
2029.360	221.879	18561.055
335.766	37.362	3017.455
282.931	29.605	2703.951
398.707	38.605	4117.788
76.488	8.353	700.410
479.841	51.959	4431.279
1189.618	129.615	10918.466
35.969	3.599	359.456
102.212	11.071	943.698
2063.614	194.446	21900.684
49.178	5.420	446.180
146.880	16.230	1329.218
216.623	21.603	2172.134
562.025	54.941	5749.239
6725.209	698.468	64753.751

fit	lwr	upr
109.608	11.751	1022.350
746.057	67.439	8253.387
72.407	7.878	665.537
46.702	4.832	451.411
4293.282	130.853	140862.481
55.624	5.865	527.551
201.277	22.397	1808.827
180.600	19.199	1698.836
3.794	0.155	93.120
454.639	48.357	4274.423
95.801	10.270	893.645
170.075	17.263	1675.602
2968.981	322.855	27302.808
2423.679	264.721	22190.232
74.181	7.775	707.734
47.433	4.568	492.544
35.739	3.822	334.200
41.227	4.194	405.244
3348.528	352.993	31764.482
49.126	5.124	471.010
19.683	2.165	178.977
169.150	18.677	1531.941
358.723	36.388	3536.344
409.339	45.415	3689.488
7302.495	733.723	72679.289
10076.000	1091.022	93055.686
1818.534	192.576	17172.747
199.682	21.474	1856.821
34.270	3.502	335.376
24.617	2.366	256.121
197.441	20.436	1907.535
560.856	53.167	5916.441
60.626	6.368	577.193
48.867	4.798	497.713
46.475	4.807	449.325
32.561	3.508	302.247
426.103	45.231	4014.191
165.149	16.950	1609.103
461.114	48.952	4343.583
61.508	5.778	654.705
75.301	8.107	699.400
831.224	85.387	8091.797
16.387	1.749	153.514
830.149	86.266	7988.638
1217.035	133.758	11073.520
68.823	7.521	629.764
101.151	9.954	1027.843
2173.597	229.075	20624.391
440.055	46.637	4152.258
1397.065	152.955	12760.539
295.563	32.951	2651.133
172.245	18.436	1609.267

fit	lwr	upr
113.694	12.193	1060.174
87.419	9.612	795.036
17.838	1.098	289.802
171.459	19.026	1545.174
18.578	1.963	175.806
14.639	1.430	149.855
132.283	14.485	1208.085
233.957	25.070	2183.307
90.932	9.946	831.343
61.882	3.239	1182.136
1430.445	152.248	13439.717
4182.932	452.551	38662.828
367.715	35.558	3802.630
305.793	34.100	2742.207
264.843	29.322	2392.133
496.363	54.633	4509.670
140.232	15.511	1267.834
142.103	13.031	1549.576
115.199	11.816	1123.120
520.153	51.094	5295.357
2706.192	291.820	25095.906
114.831	10.308	1279.172
30.915	3.282	291.176
340.750	35.784	3244.732
966.760	104.066	8981.057
28.176	2.751	288.569
24.126	2.569	226.547
442.320	48.788	4010.115
261.190	27.249	2503.612
230.358	24.737	2145.161
59.438	6.149	574.530
52.987	5.858	479.253
316.445	35.155	2848.488
21.808	2.133	222.985
99.683	10.235	970.852
220.705	21.640	2250.973
4.905	0.534	45.019
85.486	9.424	775.488
54.049	5.269	554.443
309.679	30.495	3144.809
2527.133	145.976	43749.558
39.300	4.272	361.513
4100.864	419.024	40133.997
342.676	36.675	3201.818
25.204	2.716	233.932
549.783	53.777	5620.654
40.894	4.382	381.635
7401.133	717.252	76370.281
147.975	11.271	1942.726
1008.051	108.520	9363.875
100.465	10.054	1003.900
1111.299	114.338	10801.221

fit	lwr	upr
75.391	8.028	708.034
87.248	8.577	887.512
165.160	18.233	1496.039
241.113	25.006	2324.814
140.693	15.402	1285.219
17.457	1.897	160.655
48.924	5.361	446.499
157.989	17.409	1433.737
4856.825	433.229	54448.671
280.928	29.472	2677.840
98.129	10.308	934.140
291.623	28.875	2945.198
157.740	16.846	1477.052
345.540	25.516	4679.371
139.163	13.559	1428.302
216.221	21.719	2152.603
15.616	1.597	152.668
1850.755	194.171	17640.603
29.639	3.165	277.569
55.110	6.065	500.743
53.564	5.653	507.579
73.485	7.819	690.671
21.329	2.144	212.217
241.200	25.438	2287.004
605.289	63.747	5747.374
66.199	7.239	605.389
1715.905	165.117	17831.813
46.702	4.832	451.411
53.103	5.825	484.118
47.540	5.160	437.994
37.672	4.088	347.178
562.904	61.197	5177.714
100.465	10.054	1003.900
57.105	6.112	533.539
39.094	4.060	376.444
349.541	32.559	3752.546
5.365	0.490	58.740
17.514	1.880	163.206
244.142	25.742	2315.476
3222.875	300.281	34590.622
38.635	3.938	379.029
188.766	20.348	1751.137
23.891	2.609	218.762
635.958	55.964	7226.779
115.369	12.041	1105.380
9414.470	944.877	93802.888
52.704	5.524	502.818
764.674	50.471	11585.379
2412.257	240.003	24245.484
123.472	13.181	1156.588
178.913	19.403	1649.709
814.485	86.622	7658.428

fit	lwr	upr
317.586	34.572	2917.460
207.554	21.735	1981.986
2251.886	243.641	20813.384
8467.845	921.104	77846.129
630.912	67.575	5890.510
20.731	2.115	203.223
281.230	30.573	2586.907
24.071	2.563	226.042
13.789	1.484	128.090
298.652	33.298	2678.614
62.331	6.695	580.330
88.698	8.882	885.784
863.989	94.753	7878.149
678.777	73.202	6294.030
99.026	10.963	894.479
115.176	12.675	1046.619
203.747	21.486	1932.078
502.019	53.017	4753.593
1916.507	208.914	17581.431
28.974	3.127	268.467
845.497	91.652	7799.812
25.867	2.667	250.869
612.714	64.830	5790.826
75.487	7.206	790.775
133.034	14.097	1255.413
14.994	1.425	157.756
369.362	39.518	3452.343
430.367	38.300	4835.887
7.067	0.712	70.111
4.759	0.496	45.702
74.912	7.894	710.893
149.989	14.743	1525.965
3184.708	344.306	29457.443
148.259	15.576	1411.217
72.342	7.391	708.023
19.035	1.808	200.425
226.236	25.142	2035.782
39.725	4.341	363.566
90.310	9.541	854.863
285.675	30.614	2665.747
649.843	70.715	5971.809
358.834	36.394	3538.031
389.440	40.447	3749.675
14.810	1.599	137.188
39.518	4.223	369.788
55.166	5.980	508.888
143.342	14.935	1375.761
1566.647	171.222	14334.496
40.477	4.408	371.665
350.971	38.147	3229.142
7313.951	723.258	73962.340
93.877	9.187	959.273

fit	lwr	upr
27.929	2.789	279.714
555.297	51.955	5935.020
369.051	41.092	3314.470
53.406	5.688	501.449
5417.686	583.786	50277.545
63.701	6.434	630.663
44.892	4.058	496.676
100.073	11.054	905.975
96.450	10.394	894.966
1135.825	115.208	11197.949
960.973	83.942	11001.246
212.548	23.003	1963.900
174.842	17.491	1747.763
1934.808	176.267	21237.537
145.504	15.711	1347.548
10.582	1.148	97.542
1009.640	108.813	9368.156
147.861	16.178	1351.369
56.877	5.995	539.649
266.159	29.136	2431.412
122.731	13.301	1132.470
69.871	7.728	631.751
116.039	10.025	1343.176
9.391	1.008	87.460
99.385	10.038	984.008
1136.673	122.584	10539.940
432.086	47.151	3959.590
1593.938	149.088	17041.239
7.067	0.712	70.111
1024.894	110.864	9474.741
164.906	18.329	1483.639
470.102	52.256	4229.096
923.266	79.356	10741.740
51.452	5.662	467.554
459.035	49.397	4265.704
80.722	8.787	741.597
189.436	19.598	1831.058
127.486	13.539	1200.485
554.489	60.142	5112.171
754.403	83.247	6836.584
3.946	0.418	37.212
146.073	16.114	1324.170
148.293	16.360	1344.152
1625.748	163.368	16178.505
43.817	4.560	420.993
480.654	51.453	4490.091
85.483	9.074	805.299
211.065	20.293	2195.285
298.400	33.004	2697.929
149.989	14.743	1525.965
330.995	32.187	3403.786
258.188	28.572	2333.107

fit	lwr	upr
16.747	1.754	159.913
329.821	36.549	2976.298
1717.173	175.524	16799.356
26.186	2.629	260.803
96.878	10.514	892.622
530.211	51.143	5496.835
64.633	7.160	583.415
2424.537	265.415	22147.916
38.202	4.022	362.890
117.592	12.261	1127.767
66.363	7.349	599.273
60.092	6.622	545.309
130.443	7.144	2381.917
118.087	13.007	1072.082
226.471	25.239	2032.129
36.544	3.937	339.233
19.602	1.684	228.184
546.504	53.655	5566.442
205.881	21.365	1983.969
3129.044	314.552	31126.541
92.171	10.025	847.422
234.003	23.527	2327.394
10.241	1.100	95.371
93.238	8.978	968.245
44.188	4.326	451.374
226.270	24.708	2072.157
73.021	7.735	689.310
81.272	8.966	736.691
130.805	13.331	1283.421
38.558	3.977	373.781
5.928	0.611	57.526
23.030	2.322	228.432
633.446	66.719	6014.119
966.760	104.066	8981.057
153.524	15.059	1565.110
424.514	43.779	4116.426
296.653	31.445	2798.617
15.974	1.695	150.521
98.385	7.991	1211.321
3.886	0.365	41.322
9.278	1.005	85.613
74.057	8.086	678.234
34.568	3.704	322.589
425.845	46.294	3917.182
32.456	3.557	296.162
165.287	17.480	1562.940
43.550	4.474	423.901
17.319	1.830	163.926
3003.452	297.401	30331.893
1085.917	118.738	9931.239
38.849	3.730	404.647
164.272	16.443	1641.101

fit	lwr	upr
1576.231	147.418	16853.462
30.867	3.285	290.011
99.026	10.963	894.479
27.567	2.765	274.812
118.626	12.961	1085.693
632.097	59.982	6661.164
5.510	0.584	51.968
65.171	7.164	592.829
106.252	11.266	1002.048
1328.719	141.372	12488.324
40.585	4.325	380.845
13.871	1.347	142.839
245.121	27.228	2206.678
57.734	5.801	574.639
92.633	7.513	1142.154
742.131	80.848	6812.276
44.184	4.872	400.682
242.326	25.977	2260.554
150.293	16.610	1359.914
134.707	13.106	1384.551
197.441	20.436	1907.535
112.026	12.054	1041.098
28.509	2.618	310.417
9.087	0.984	83.872
39.518	4.223	369.788
7.181	0.623	82.722
41.617	4.505	384.433
104.888	11.624	946.457
271.647	29.872	2470.264
86.077	7.945	932.585
45.042	4.260	476.213
14.066	1.470	134.611
46.702	4.832	451.411
767.826	84.030	7016.016
23.578	2.582	215.307
224.968	24.535	2062.770
84.287	9.110	779.828
3.817	0.156	93.667
223.430	24.503	2037.356
2931.279	267.708	32096.096
386.470	39.858	3747.258
12.151	1.301	113.524
664.885	54.421	8123.236
5361.758	565.086	50874.502
55.241	5.987	509.657
4.038	0.408	40.009
80.105	8.656	741.288
59.981	4.886	736.258
758.447	60.975	9434.057
525.298	54.001	5109.883
633.446	66.719	6014.119
472.110	50.728	4393.817

fit	lwr	upr
383.045	24.812	5913.369
596.248	54.726	6496.161
191.625	21.270	1726.356
63.634	7.051	574.297
237.767	21.214	2664.860
981.419	100.621	9572.414
96.450	10.394	894.966
33.684	3.549	319.722
5622.068	332.970	94926.404
73.801	8.029	678.384
51.678	5.314	502.525
124.276	13.457	1147.659
43.594	4.434	428.584
64.633	7.160	583.415
78.182	8.382	729.214
1115.057	96.998	12818.275
677.712	73.355	6261.221
1640.547	163.182	16493.167
33.965	3.553	324.719
122.703	13.560	1110.321
73.513	8.075	669.252
270.822	29.659	2472.926
378.220	42.045	3402.280
39.094	4.060	376.444
62.083	6.880	560.214
616.539	58.619	6484.628
59.788	6.626	539.527
251.949	25.519	2487.534
4342.120	435.696	43273.308
32.561	3.508	302.247
10.860	1.047	112.631
4016.128	411.959	39152.605
94.990	8.790	1026.537
98.172	8.727	1104.385
317.004	28.601	3513.532
1235.113	134.521	11340.278
57.175	5.988	545.969
145.544	15.974	1326.071
84.129	9.036	783.289
2.006	0.098	41.041
342.190	36.014	3251.369
143.651	14.043	1469.468
877.035	95.870	8023.297
386.077	42.538	3504.067
153.140	16.936	1384.741
782.874	64.835	9453.022
39.094	4.060	376.444
9.469	1.007	89.053
1471.267	153.378	14113.040
45.137	4.907	415.200
7.525	0.783	72.306
208.159	22.638	1914.072

fit	lwr	upr
203.815	22.067	1882.517
764.057	75.538	7728.388
186.739	20.687	1685.692
633.446	66.719	6014.119
150.877	16.031	1419.977
854.482	92.990	7851.833
105.728	10.134	1103.053
29.776	3.228	274.702
1322.642	113.217	15451.621
190.729	20.230	1798.181
77.833	6.951	871.537
2237.332	216.577	23112.528
903.429	95.942	8507.092
63.564	6.377	633.541
25.896	2.700	248.419
36.039	3.342	388.675
14.679	1.412	152.564
42.772	4.680	390.872
208.822	22.680	1922.727
1669.918	166.717	16726.689
127.319	13.153	1232.483
28.296	2.785	287.448
123.823	13.661	1122.364
663.848	70.910	6214.804
67.923	7.091	650.591
8.900	0.964	82.175
1717.344	164.106	17971.775
37.748	4.150	343.378
18.342	1.719	195.721
183.561	20.333	1657.103
216.519	22.858	2050.943
481.768	41.344	5613.860
29.779	2.952	300.372
156.531	16.255	1507.325
117.306	12.448	1105.477
176.758	18.735	1667.691
70.243	6.591	748.618
312.431	32.866	2970.017
417.977	46.279	3775.054
145.462	15.344	1378.988
190.744	21.035	1729.682
158.933	15.308	1650.151
27.709	2.904	264.388
914.530	98.828	8462.790
66.715	6.906	644.475
12889.890	727.948	228243.311
240.990	26.416	2198.559
24.617	2.366	256.121
167.480	16.562	1693.643
58.441	3.526	968.695
922.587	95.537	8909.279
190.921	19.874	1834.046

fit	lwr	upr
56.006	4.704	666.852
343.792	38.168	3096.624
76.334	7.762	750.680
173.512	19.187	1569.148
39.094	4.060	376.444
89.563	7.454	1076.070
142.624	14.860	1368.879
93.002	9.299	930.164
32.076	3.380	304.411
225.778	24.627	2069.875
39.094	4.060	376.444
78.251	8.289	738.746
64.698	7.099	589.668
1109.350	111.647	11022.737
149.583	16.421	1362.575
41.477	4.384	392.429
2860.326	310.587	26341.918
1318.784	144.118	12067.843
520.429	37.582	7206.869
3010.222	270.930	33445.663
131.218	13.167	1307.710
1642.869	179.638	15024.744
521.200	57.268	4743.451
614.098	65.244	5780.069
1808.500	196.379	16654.911
163.011	17.590	1510.662
71.473	6.836	747.232
1627.124	65.987	40121.878
59.700	6.565	542.915
63.676	6.983	580.681
9.372	1.016	86.475
165.149	16.950	1609.103
431.312	44.541	4176.638
33.918	3.702	310.768
21.735	2.240	210.912
300.638	31.047	2911.180
293.952	32.121	2690.036
119.558	12.296	1162.508
257.457	28.491	2326.529
3093.886	295.737	32367.010
15.320	1.598	146.878
55.631	5.886	525.824
61.517	6.265	604.011
663.848	70.910	6214.804
157.830	17.464	1426.414
39.537	4.317	362.085
480.336	49.924	4621.439
66.927	7.204	621.764
495.569	54.545	4502.532
852.825	92.809	7836.666
328.477	33.033	3266.294
10.737	1.118	103.110

fit	lwr	upr
40.948	4.133	405.706
362.673	37.933	3467.516
58.350	5.943	572.862
41.106	4.228	399.662
1499.320	155.554	14451.328
62.712	6.896	570.327
589.873	63.207	5504.964
597.252	66.244	5384.797
205.491	22.604	1868.069
329.760	36.577	2972.970
312.605	33.428	2923.376
316.392	32.390	3090.590
36.400	3.953	335.226
2078.903	170.183	25395.257
85.593	8.473	864.648
378.776	41.602	3448.682
344.823	29.646	4010.804
78.785	7.759	800.008
219.213	23.664	2030.682
41.617	4.505	384.433
117.597	12.979	1065.478
36.783	3.616	374.201
5.977	0.641	55.694
289.579	29.294	2862.556
713.973	68.187	7475.842
161.173	17.729	1465.192
2.560	0.138	47.470
32.364	3.370	310.850
104.787	10.989	999.236
112.906	10.592	1203.497
248.313	27.586	2235.146
126.989	12.032	1340.296
86.561	9.536	785.753
129.114	14.356	1161.209
65.536	6.863	625.848
92.736	8.653	993.822
26.112	2.756	247.439
1160.114	127.314	10571.222
34.568	3.704	322.589
90.310	9.541	854.863
222.490	24.104	2053.633
45.308	4.730	434.002
62.083	6.880	560.214
3.039	0.287	32.208
69.629	7.702	629.501
80.628	8.295	783.687
8410.264	740.990	95456.807
33.312	3.506	316.531
51.776	4.934	543.277
132.102	14.067	1240.555
47.315	5.178	432.380
4409.059	478.822	40599.201

fit	lwr	upr
271.576	30.285	2435.336
20.862	0.950	457.919
947.047	83.952	10683.476
369.040	40.815	3336.743
38.149	4.087	356.095
30.944	2.930	326.863
23.115	1.967	271.594
59.875	5.193	690.402
97.219	10.053	940.168
116.219	11.629	1161.479
40.585	4.325	380.845
76.691	8.065	729.277
24.961	2.692	231.441
506.542	52.171	4918.122
284.955	20.844	3895.592
85.171	8.954	810.148
112.243	12.368	1018.610
17.818	1.914	165.900
50.512	4.987	511.584
37.224	4.069	340.491
66.707	7.334	606.713
23.605	2.585	215.552
580.246	55.743	6039.969
120.515	12.459	1165.763
123.113	13.607	1113.900
672.532	74.157	6099.191
742.411	76.224	7231.004
7.940	0.855	73.771
301.046	30.233	2997.634
45.135	1.698	1199.652
258.473	28.604	2335.670
187.763	19.615	1797.345
298.016	26.391	3365.312
19.656	1.820	212.299
253.057	26.405	2425.196
304.685	33.839	2743.362
43.288	4.321	433.669
169.983	16.228	1780.477
562.025	54.941	5749.239
244.142	25.742	2315.476
104.668	11.114	985.726
25.916	2.584	259.938
1636.749	163.362	16398.864
240.065	24.508	2351.501
87.904	9.711	795.754
96.712	10.288	909.102
108.268	10.533	1112.915
884.196	91.786	8517.677
41.269	4.107	414.669
3098.906	327.090	29359.556
280.651	28.548	2759.010
106.686	11.617	979.760

fit	lwr	upr
198.903	19.506	2028.216
454.181	48.699	4235.813
241.328	26.326	2212.208
247.868	27.051	2271.163
86.683	9.153	820.959
262.991	27.067	2555.284
7.822	0.721	84.816
369.782	39.990	3419.297
167.480	16.562	1693.643
499.557	54.296	4596.234
247.425	27.487	2227.221
98.285	10.754	898.255
2937.183	319.817	26974.964
226.270	24.708	2072.157
76.719	8.452	696.345
7.119	0.742	68.341
40.880	2.794	598.134
270.625	29.413	2490.012
536.272	54.742	5253.487
560.490	60.462	5195.820
51.682	5.485	487.001
230.850	24.846	2144.855
1043.595	97.644	11153.709
531.906	45.680	6193.613
131.218	13.167	1307.710
1576.231	147.418	16853.462
102.092	10.652	978.460
2.002	0.179	22.379
13.614	1.398	132.612
15.727	1.540	160.640
241.328	26.326	2212.208
43.705	4.600	415.243
417.734	45.295	3852.517
60.746	6.573	561.412
81.371	8.934	741.119
336.693	37.466	3025.704
2605.402	262.643	25845.429
56.579	5.850	547.223
109.662	12.077	995.717
514.621	43.376	6105.508
314.910	28.746	3449.829
29.392	3.167	272.781

Since we are still using the linear regression model, we used `predict` function with `interval = "pred"` argument to obtain the prediction interval.

5. Assessment of the final model (25 points)

Model evaluation: must include an evaluation discussion

Model testing : must include a discussion

Model result: must include a selection and discussion of the top 10 valued paintings in the validation data.

6. Conclusion (10 points): must include a summary of results and a discussion of things learned. Optional what would you do if you had more time.