

STA 521 - Final Project Part I

FP-Team 01: Qianyin Lu, George Lindner, Chenxi Wu, Yi Mi

December 7th, 2019

Write-Up

1. Introduction: Summary of problem and objectives

Our team of esteemed statisticians was recently hired by a prestigious Art historian for a consulting project. We were asked to help build a predictive model in exchange for an A on our STA 521 Final Exam. After much discussion, our team accepted the historian's offer.

We were given the task of predicting paintings' selling prices at auctions in 18th century Paris. To accomplish this, we used a dataset containing information about each painting's buyer, seller, painter, and characteristics of the painting. These variables were all possible predictor variables in modeling the response variable, the selling price of a painting.

There were two primary objectives in our analysis:

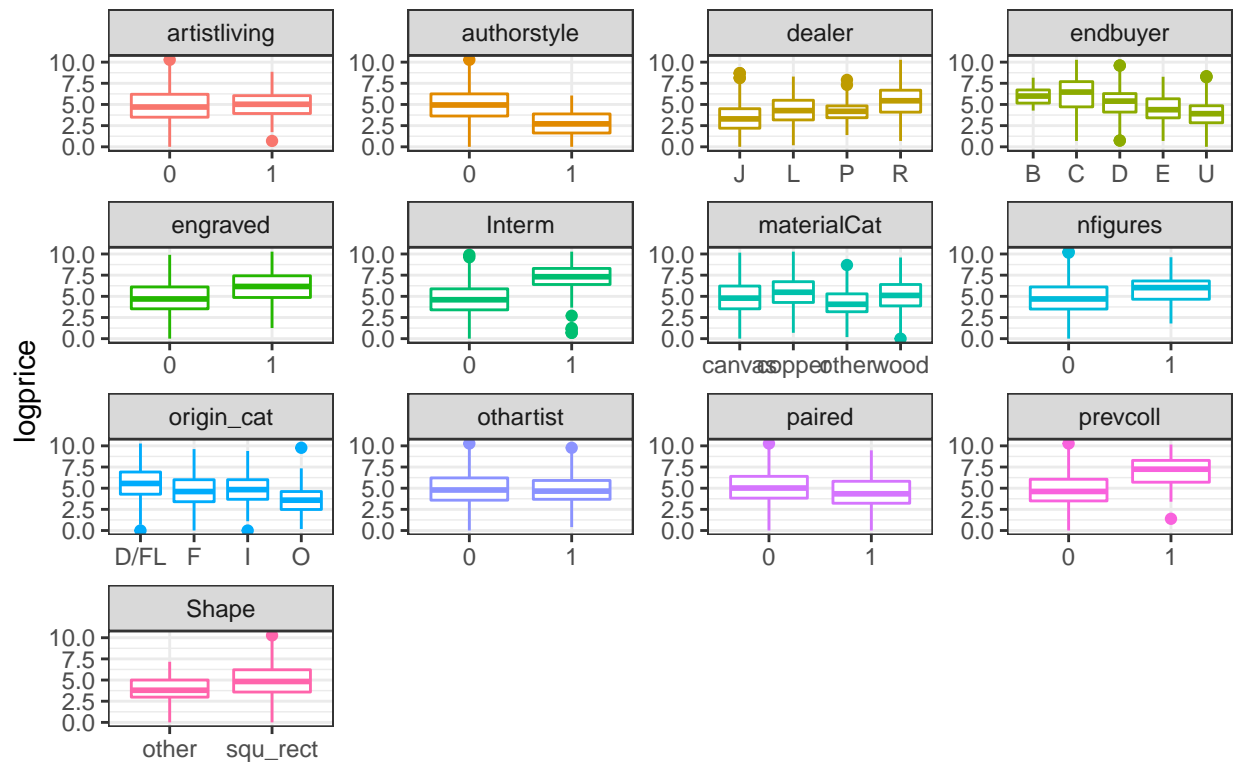
- 1) To determine which variables (or interactions) drove the price of a painting
- 2) To determine which paintings were overpriced or and which were underpriced.

After arriving at a final model, we are able to answer these primary questions. Any variables that appear in the model will be important in driving painting prices, and observing residuals will enable us to determine if a painting was over or underpriced.

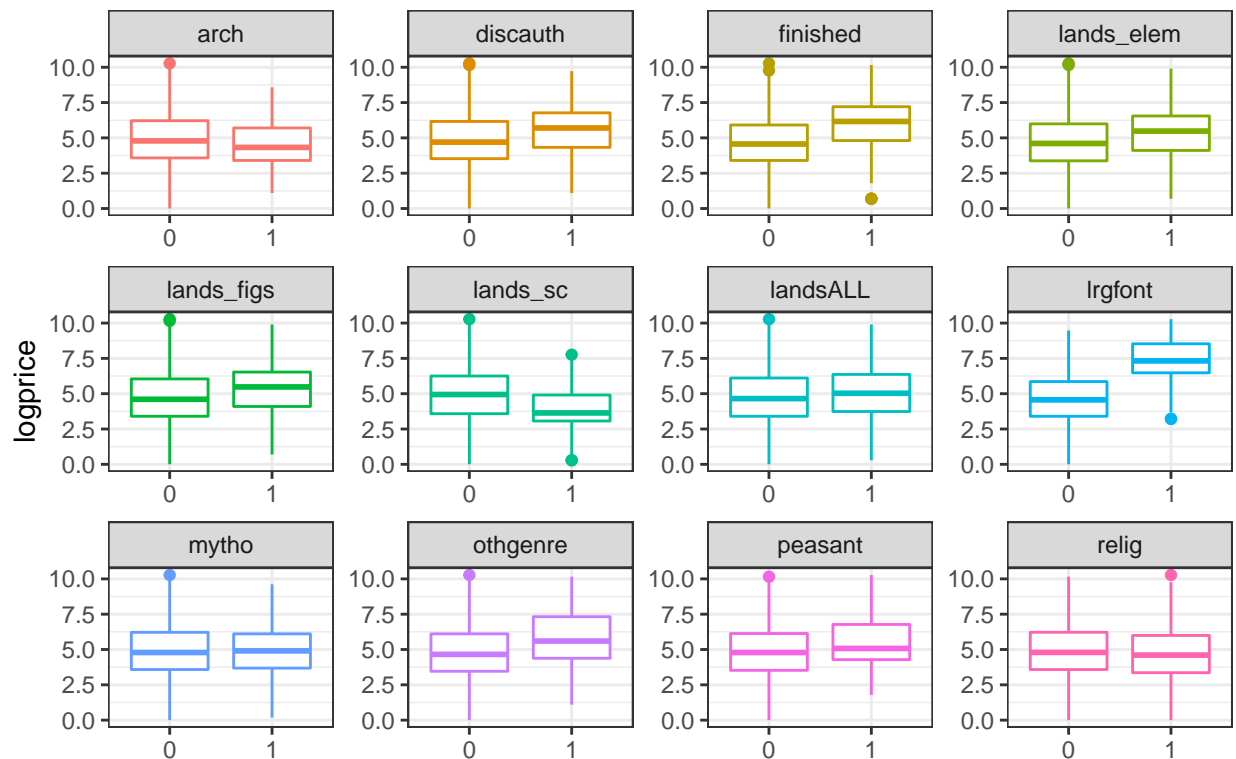
We had 1,500 observations to train the model on, along with 750 observations held out as a testing set. There were a total of 59 variables in the dataset, both categorical and continuous.

2. Exploratory Data Analysis:

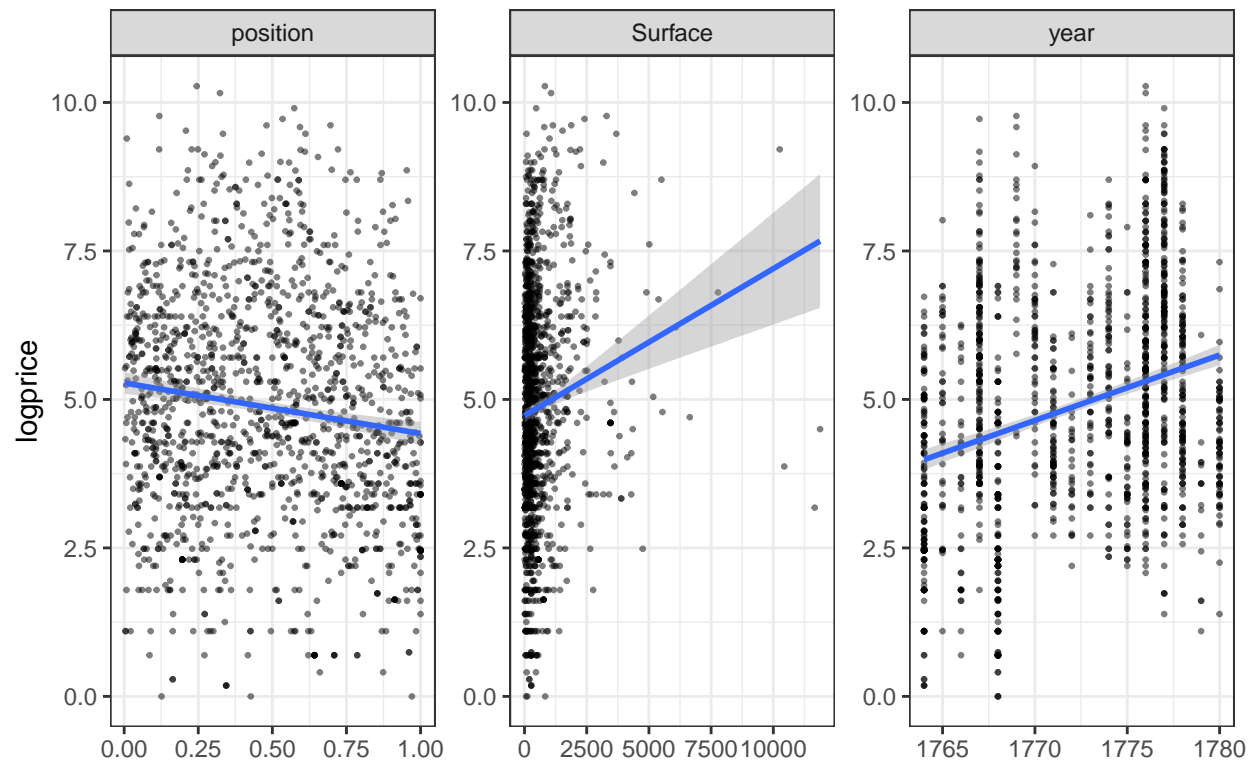
Log Price vs Categorical Predictors



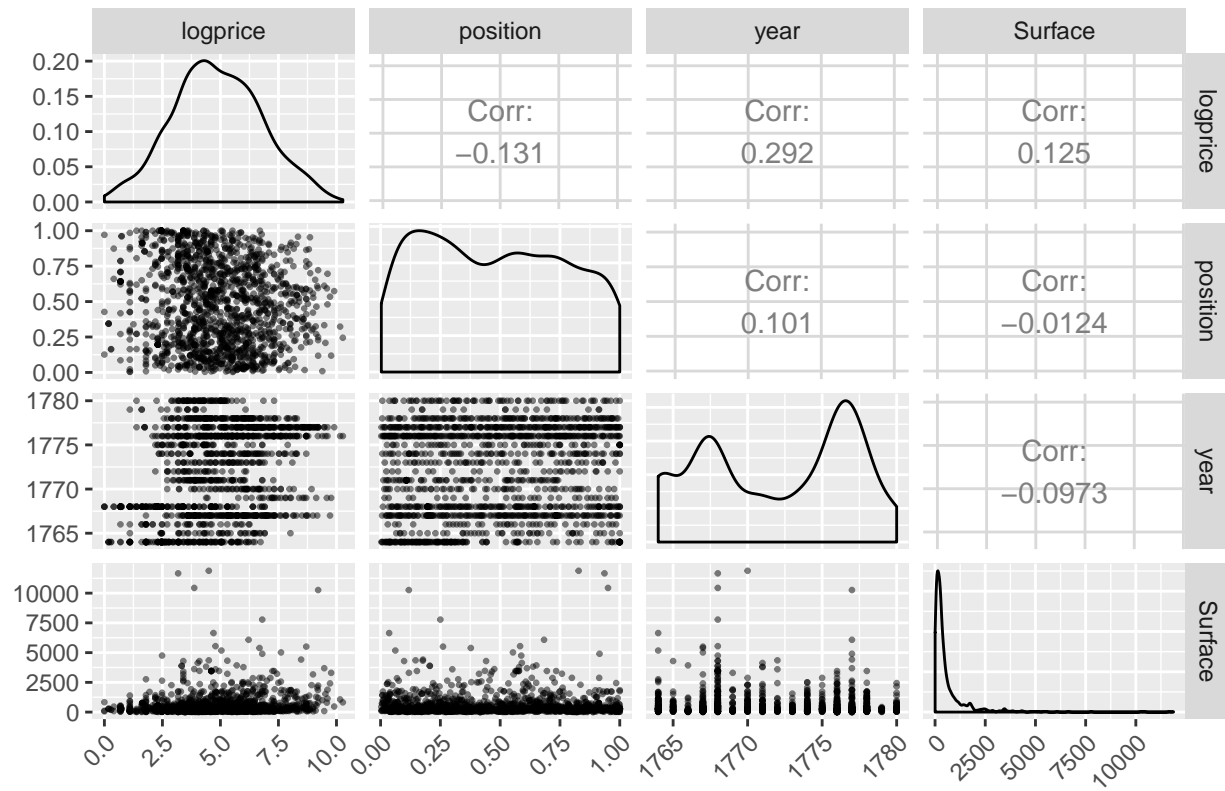
Log Price vs Categorical Predictors



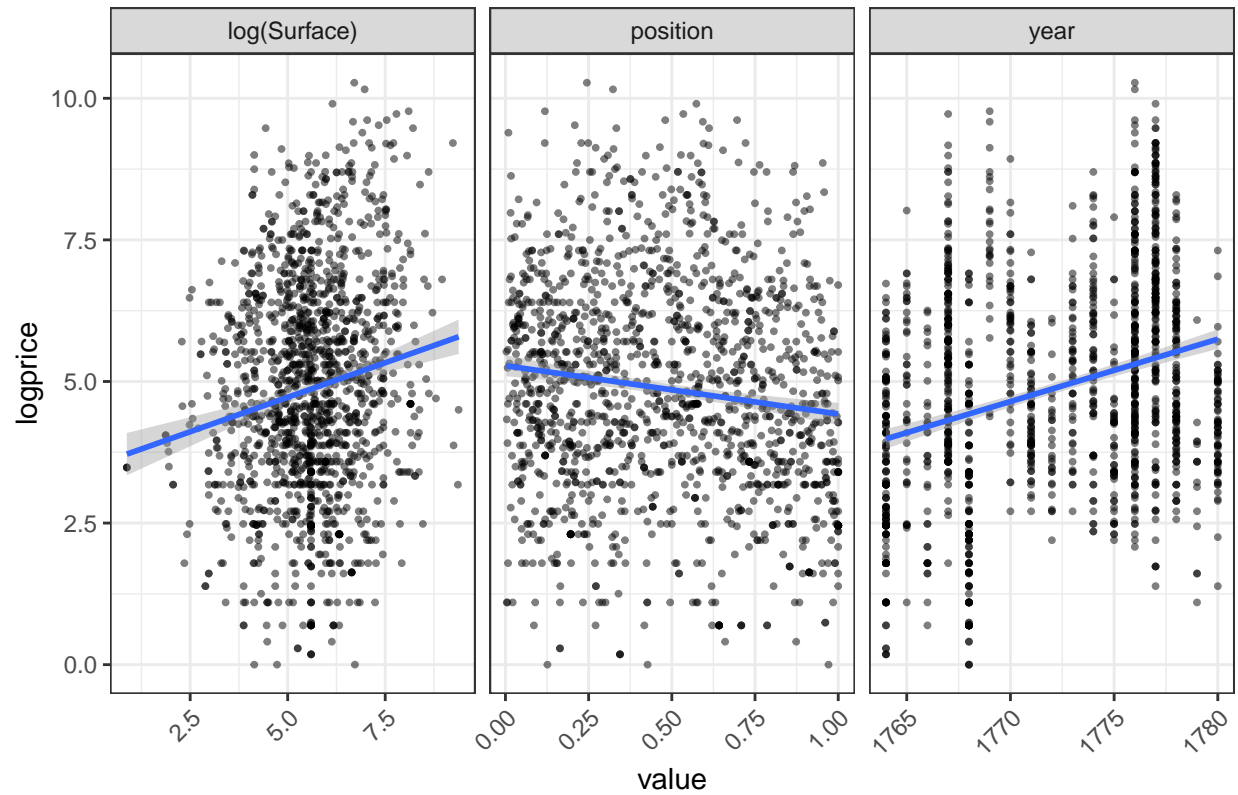
Log Price vs Continuous Predictors



Pairwise Comparisons of Continuous Variables



Log Price vs Continuous Predictors



- from the EDA plot, decided to further drop *othartist*, *mytho*, *relig* and *landsALL*

3. Development and Assessment of Initial Model:

4. Summary and Conclusions: