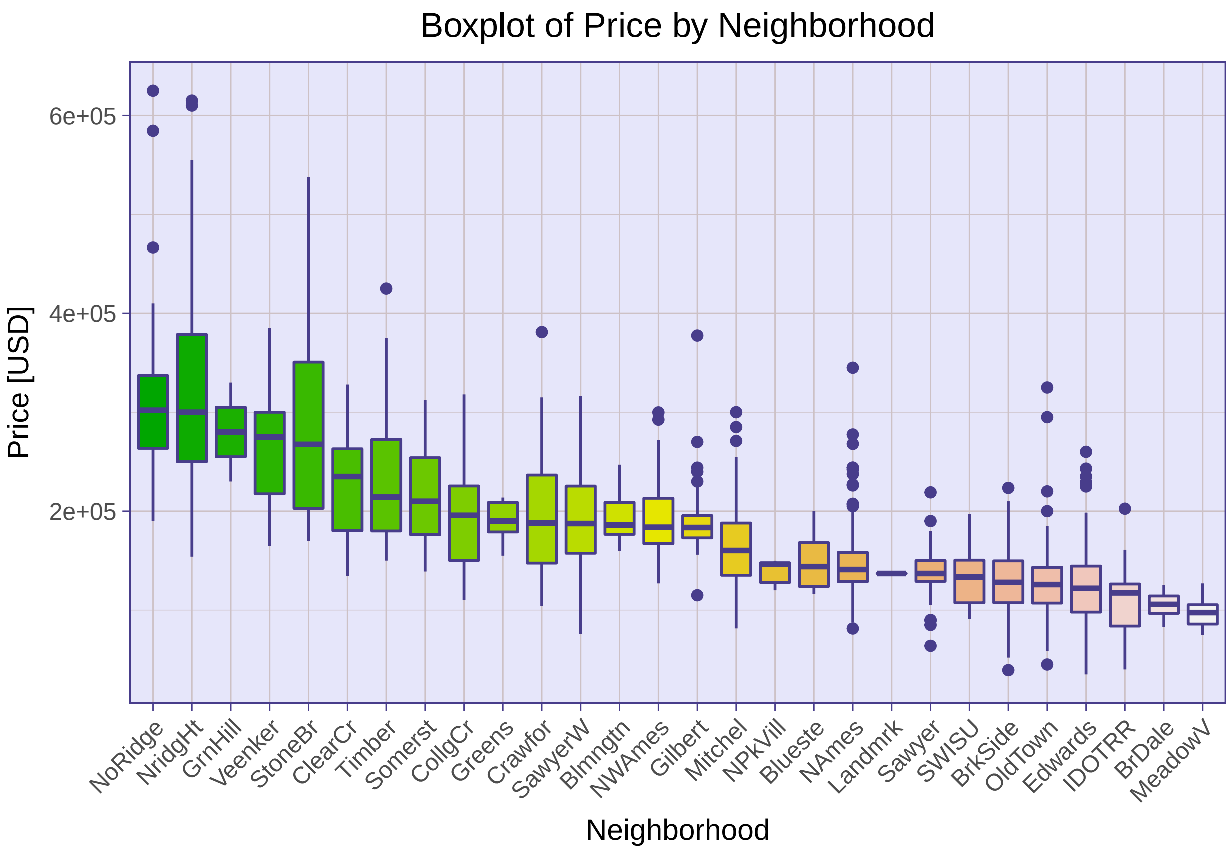
Final Project Writeup

Thomas Fleming, Eden Huang, Blaire Li, Marc Ryser

1. Exploratory data analysis (20 points)

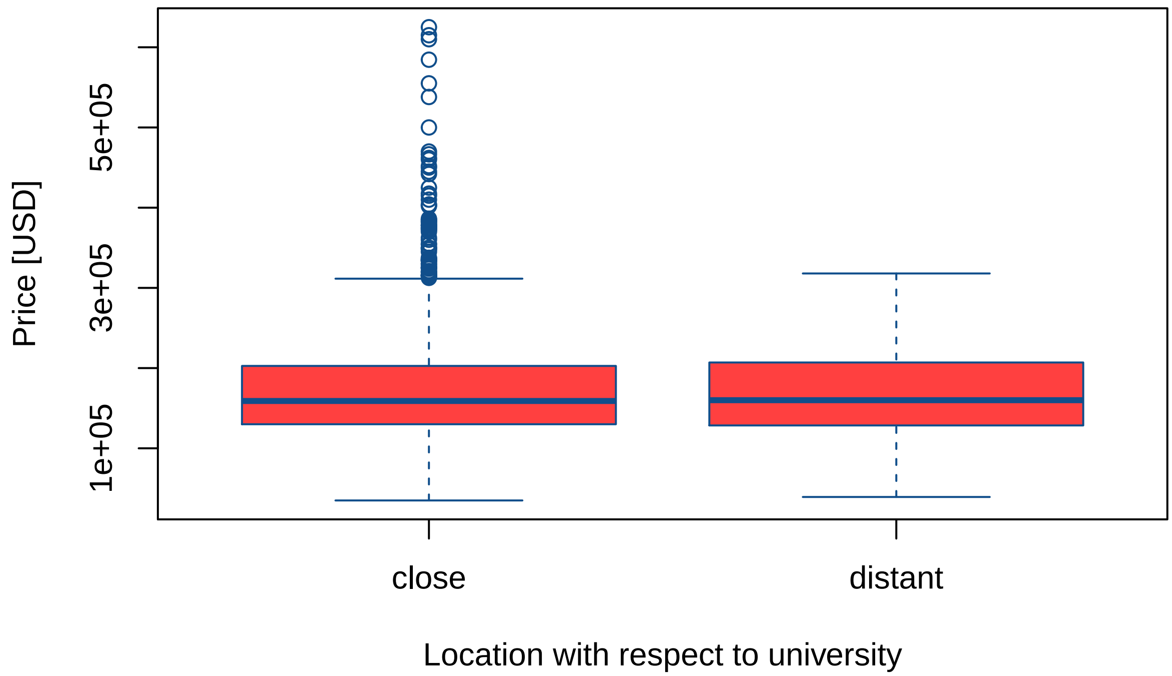
We first performed basic checks on the training data to identify predictors with (truly) missing entries and potential abnormalities. The variable lot frontage was identified to have 282 or 18.8% missing entries. We removed this predictor, and verified at a later stage that adding it in would not lead to better model predictions.

The mantra in real estate appears to be "location, location, location." This begs for a simple visualization of house price distribution by neighborhood.

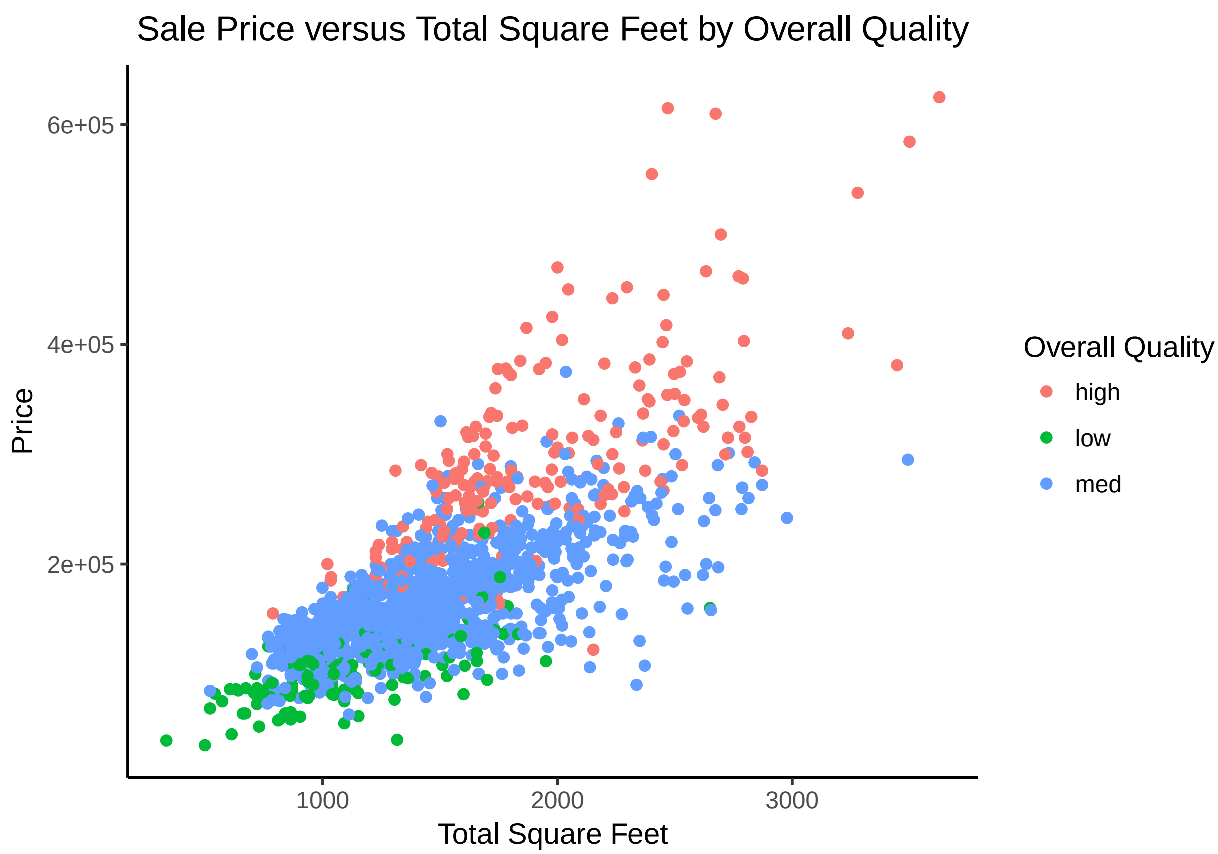


An interesting observation is that there is a wider dispersion among the more affluent neighborhoods based on relative inter-quartile ranges, whereas the neighborhoods with cheaper housing tend to be more concentrated around their medians. In other words, homoscedasticity is violated in this data set as variation increases with sale price. This is an important observation, as it indicates that we would be wise to transform the response variable (which we eventually did).

With respect to location, we note that Ames is a college town built around Iowa State. With almost 16,000 employees, the university is by far the biggest employer in town. Using Google maps, we recorded the respective distances between the university to the different neighborhoods. Beyond a univariable analysis, we could not find an association between price and distance to university. This is illustrated in the following boxplot where we compare prices between close and distant houses with respect to the university (defined as within 1.5 miles to Parks Library at ISU, measured using Google Maps). However, the proximity to ISU doesn't seem to affect property prices as expected.



Our third plot shows the relationship between sales price and total square footage, stratified by overall quality of the house. We see that there are different slopes in the relationship between sales price and total square footage, indicating the need for interaction terms between these variables (indeed, the more complex model contains such interactions).



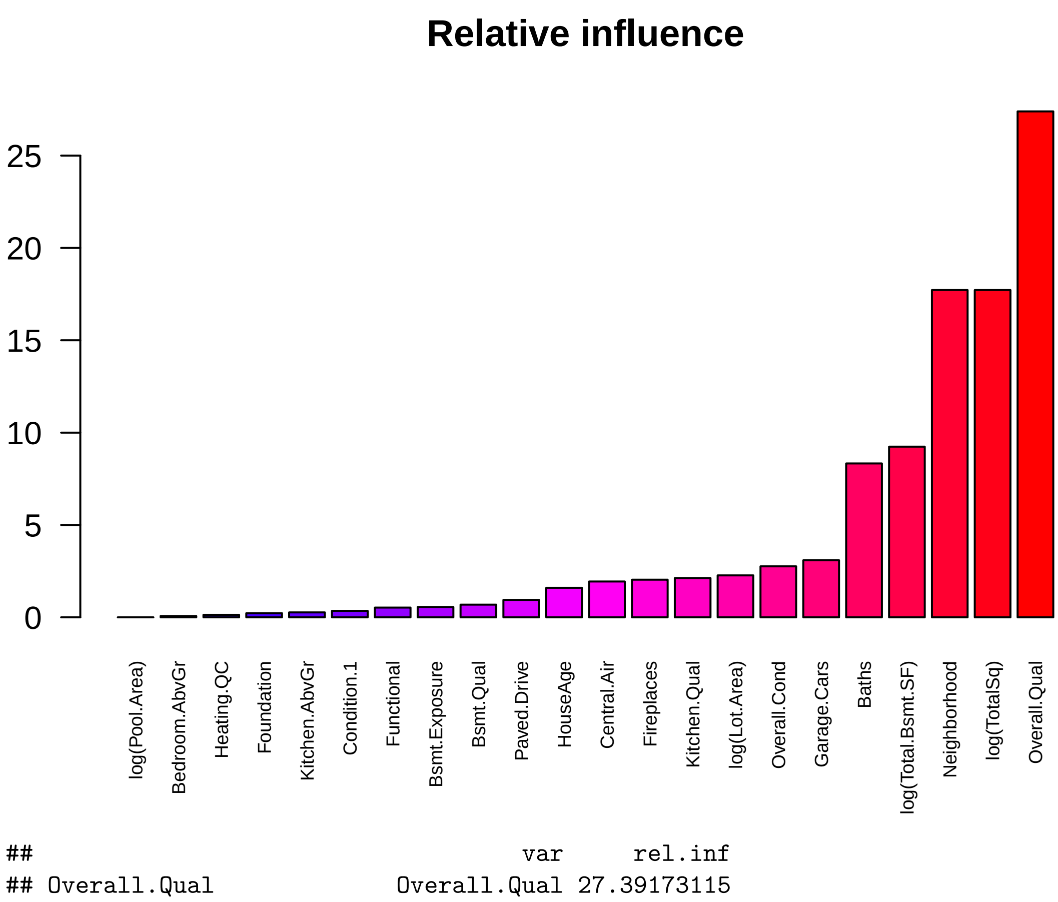
2. Development and assessment of an initial model from Part I (10 points)

Initial Model and Model Selection

Prior to selecting the variables for our simple model, we cleaned the data (big time). We used a variety of techniques to do this, including converting variables to factors to account for non-linearities (e.g. MS.Subclass, Alley, Bsmt.Qual, Bsmt.Cond, etc.), aggregating like variables, such as combining porch square footage for different types of porches and creating a variable for total number of baths, accounting for the NA's by adding levels named "None" where appropriate, creating a variable for house age- calculated by subtracting the max of year built and year remodeled from the year sold- and filtering out NA's for a few particular variables with few NA's.

Following our data cleaning, we performed exploratory analyses. We first fit a full linear model, including all available variables, and examined its residual plots for indication as to the changes we should make. We detected some non-linearity in the data, seeing a trend in the residual and standardized residual plots, and decided to make a log transformation of our response variable, which improved the model fit. Taking a note from Appendix A in Gelman's book, we also logged all continuous explanatory variables, as this helps provide a multiplicative effect to the model when transformed to the original scale. The BoxCox procedure also indicated that a log-transformation for the response is reasonable.

Having improved our model fit through simple variable transformations, we then went about variable selection. We found that running a step function evaluated using the Bayesian Information Criterion brought us down close to 20 predictors, giving us 22. We then used a boosted tree model (depth 1, no interactions) to evaluate the relative importance of these 22 variables, and removed log(Pool.Area), Bedrooms.AbvGr, and Heating.QC, as these variables had the lowest relative importance. This brought us to our finalized simple model, with 19 predictors.




```
##                                    var     rel.inf
## Overall.Qual             Overall.Qual 27.39173115
## log(TotalSq)             log(TotalSq) 17.71933311
```

```
## Neighborhood            Neighborhood 17.71808003
## log(Total.Bsmt.SF) log(Total.Bsmt.SF)  9.24092225
## Baths                          Baths  8.33117234
## Garage.Cars              Garage.Cars  3.09118466
## Overall.Cond            Overall.Cond  2.75985318
## log(Lot.Area)          log(Lot.Area)  2.27126022
## Kitchen.Qual            Kitchen.Qual  2.12718531
## Fireplaces                Fireplaces  2.03782711
## Central.Air              Central.Air  1.94038811
## HouseAge                    HouseAge  1.59881363
## Paved.Drive              Paved.Drive  0.94528641
## Bsmt.Qual                  Bsmt.Qual  0.68590086
## Bsmt.Exposure          Bsmt.Exposure  0.56197978
## Functional                Functional  0.52774604
## Condition.1              Condition.1  0.35115676
## Kitchen.AbvGr          Kitchen.AbvGr  0.26708845
## Foundation                Foundation  0.22452938
## Heating.QC                Heating.QC  0.13573258
## Bedroom.AbvGr          Bedroom.AbvGr  0.07282862
## log(Pool.Area)        log(Pool.Area)  0.00000000
```

```
##
## Call:
## lm(formula = log(price) ~ log(Lot.Area) + Neighborhood + Condition.1 +
##     Overall.Qual + Overall.Cond + HouseAge + Foundation + Bsmt.Qual +
##     Bsmt.Exposure + log(Total.Bsmt.SF) + Central.Air + Baths +
##     Kitchen.AbvGr + Kitchen.Qual + Functional + Fireplaces +
##     Garage.Cars + Paved.Drive + log(TotalSq), data = data_train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.66880 -0.05342 -0.00040  0.05532  0.34125
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          7.5928042  0.1034783  73.376  < 2e-16 ***
## log(Lot.Area)        0.1005736  0.0075956  13.241  < 2e-16 ***
## NeighborhoodBlueste  0.0038079  0.0422185   0.090 0.928145
## NeighborhoodBrDale  -0.0773350  0.0348923  -2.216 0.026822 *
## NeighborhoodBrkSide -0.0297326  0.0309012  -0.962 0.336120
## NeighborhoodClearCr -0.0032135  0.0340622  -0.094 0.924851
## NeighborhoodCollgCr -0.0225456  0.0276495  -0.815 0.414975
## NeighborhoodCrawfor  0.0440361  0.0313361   1.405 0.160155
## NeighborhoodEdwards -0.1090553  0.0295961  -3.685 0.000237 ***
## NeighborhoodGilbert -0.0503690  0.0287855  -1.750 0.080367 .
## NeighborhoodGreens   0.0254955  0.0444436   0.574 0.566287
## NeighborhoodGrnHill  0.4479441  0.0714376   6.270 4.76e-10 ***
## NeighborhoodIDOTRR  -0.1533929  0.0326944  -4.692 2.97e-06 ***
## NeighborhoodLandmrk -0.0763339  0.0957014  -0.798 0.425220
## NeighborhoodMeadowV -0.1420774  0.0384490  -3.695 0.000228 ***
## NeighborhoodMitchel -0.0461630  0.0296321  -1.558 0.119485
## NeighborhoodNAmes   -0.0670485  0.0286918  -2.337 0.019584 *
## NeighborhoodNoRidge  0.0541880  0.0304830   1.778 0.075674 .
## NeighborhoodNPkVill -0.0388906  0.0385797  -1.008 0.313597
```

```
## NeighborhoodNridgHt  0.0455037  0.0292245   1.557 0.119682
## NeighborhoodNWAmes  -0.0669261  0.0298203  -2.244 0.024965 *
## NeighborhoodOldTown -0.1235192  0.0296508  -4.166 3.29e-05 ***
## NeighborhoodSawyer  -0.0386395  0.0299358  -1.291 0.197000
## NeighborhoodSawyerW -0.0603434  0.0290947  -2.074 0.038255 *
## NeighborhoodSomerst  0.0500211  0.0278563   1.796 0.072757 .
## NeighborhoodStoneBr  0.0703212  0.0323876   2.171 0.030077 *
## NeighborhoodSWISU   -0.0749894  0.0341474  -2.196 0.028249 *
## NeighborhoodTimber  -0.0418053  0.0319528  -1.308 0.190967
## NeighborhoodVeenker -0.0129656  0.0378365  -0.343 0.731894
## Condition.1Artery   -0.0768610  0.0152515  -5.040 5.26e-07 ***
## Condition.1Feedr    -0.0749669  0.0114239  -6.562 7.39e-11 ***
## Condition.1Park      0.0063233  0.0183326   0.345 0.730204
## Condition.1Rail     -0.0441457  0.0149789  -2.947 0.003259 **
## Overall.Qual         0.0509388  0.0034351  14.829  < 2e-16 ***
## Overall.Cond         0.0346458  0.0027915  12.411  < 2e-16 ***
## HouseAge            -0.0009529  0.0001880  -5.069 4.52e-07 ***
## FoundationCBlock     0.0576301  0.0105526   5.461 5.57e-08 ***
## FoundationPConc      0.0749570  0.0115916   6.466 1.37e-10 ***
## FoundationSlab       0.0647669  0.0293347   2.208 0.027413 *
## FoundationStone      0.0002816  0.0393425   0.007 0.994291
## FoundationWood       0.0525257  0.0559323   0.939 0.347841
## Bsmt.QualEx         -0.7533650  0.1145897  -6.574 6.83e-11 ***
## Bsmt.QualFa         -0.8320710  0.1133422  -7.341 3.54e-13 ***
## Bsmt.QualGd         -0.8380252  0.1131896  -7.404 2.25e-13 ***
## Bsmt.QualPo         -0.8410027  0.1454982  -5.780 9.15e-09 ***
## Bsmt.QualTA         -0.8503317  0.1130739  -7.520 9.63e-14 ***
## Bsmt.ExposureAv      0.1560803  0.0926024   1.685 0.092113 .
## Bsmt.ExposureGd      0.1933630  0.0929084   2.081 0.037591 *
## Bsmt.ExposureMn      0.1117113  0.0928128   1.204 0.228936
## Bsmt.ExposureNo      0.1272957  0.0925440   1.376 0.169187
## log(Total.Bsmt.SF)   0.1196200  0.0092381  12.949  < 2e-16 ***
## Central.AirY         0.0634765  0.0119725   5.302 1.33e-07 ***
## Baths                0.0473048  0.0045589  10.376  < 2e-16 ***
## Kitchen.AbvGr       -0.1041428  0.0140674  -7.403 2.26e-13 ***
## Kitchen.Qual         0.0380916  0.0055894   6.815 1.39e-11 ***
## FunctionalMaj2      -0.1668154  0.0522781  -3.191 0.001449 **
## FunctionalMin1       0.0144202  0.0348213   0.414 0.678848
## FunctionalMin2       0.0342439  0.0344569   0.994 0.320480
## FunctionalMod       -0.0001033  0.0374934  -0.003 0.997801
## FunctionalTyp        0.0846081  0.0315133   2.685 0.007341 **
## Fireplaces           0.0313138  0.0047078   6.651 4.12e-11 ***
## Garage.Cars          0.0395438  0.0047321   8.356  < 2e-16 ***
## Paved.DriveP         0.0070829  0.0182491   0.388 0.697982
## Paved.DriveY         0.0575194  0.0115044   5.000 6.45e-07 ***
## log(TotalSq)         0.3439479  0.0130974  26.261  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09198 on 1428 degrees of freedom
## Multiple R-squared:  0.9415, Adjusted R-squared:  0.9389
## F-statistic:   359 on 64 and 1428 DF,  p-value: < 2.2e-16
```

Residual Plots

The residual plots for our simple model display favorable results. The residual vs. fitted plot shows little to no pattern and while the scale-location plot has a slight pattern, it is not severe. While there are a few outliers displayed on the Normal Q-Q plot, most observations fall within 4 standard deviations, with the majority falling very close to or on the one-to-one line. Finally, after having the function remove observations with leverage 1 (of which there were only 3), we observe a favorable leverage plot, with no observations exceeding a Cook's distance of 0.5.

RMSE

The RMSE for the simple model evaluated on the test was 15,477, which corresponds to approximately 8% of the mean house price. This implies that the RMSE is fairly small.

Model Testing

Beyond the RMSE, we were very pleased to see a bias relatively closer to zero than our fellow teams, at -165.05. We found a maximum deviation of 66,474.27, a mean deviation of 11,458.19, and coverage of 96.2%. The coverage is very favorable and, given that the mean deviation is less than 10% of the mean housing price for the training set, we are please with these results.

	bias	max.dev	mean.dev	RMSE	Coverage
Training	799.9486	152265.48	11858.23	17070.81	0.9564635
Test	-165.0479	66474.27	11458.19	15477.42	0.9620000

3. Development of the final model (20 points)

We tried a range of different approaches for a more complex model. Among others, we evaluated tree models, bagging, boosting and random forests, as well as Lasso and Ridge. Based on RMSE however, none of these options was able to outperform a linear model with interaction terms. To develop the final linear model, we proceeded as follows:

- We grew a deep true
- We pruned the tree using cross-validation
- We read out the interactions from the final tree
- We inserted these interactions into the full linear model
- We ran a stepwise BIC

The reason for not using a tree-based model for the final version was that the RMSE of these approaches remained substantially higher compared to the final OLS model.

In consequence, the final model contains 22 different variables, two 2-way interactions and one 3-way interaction.

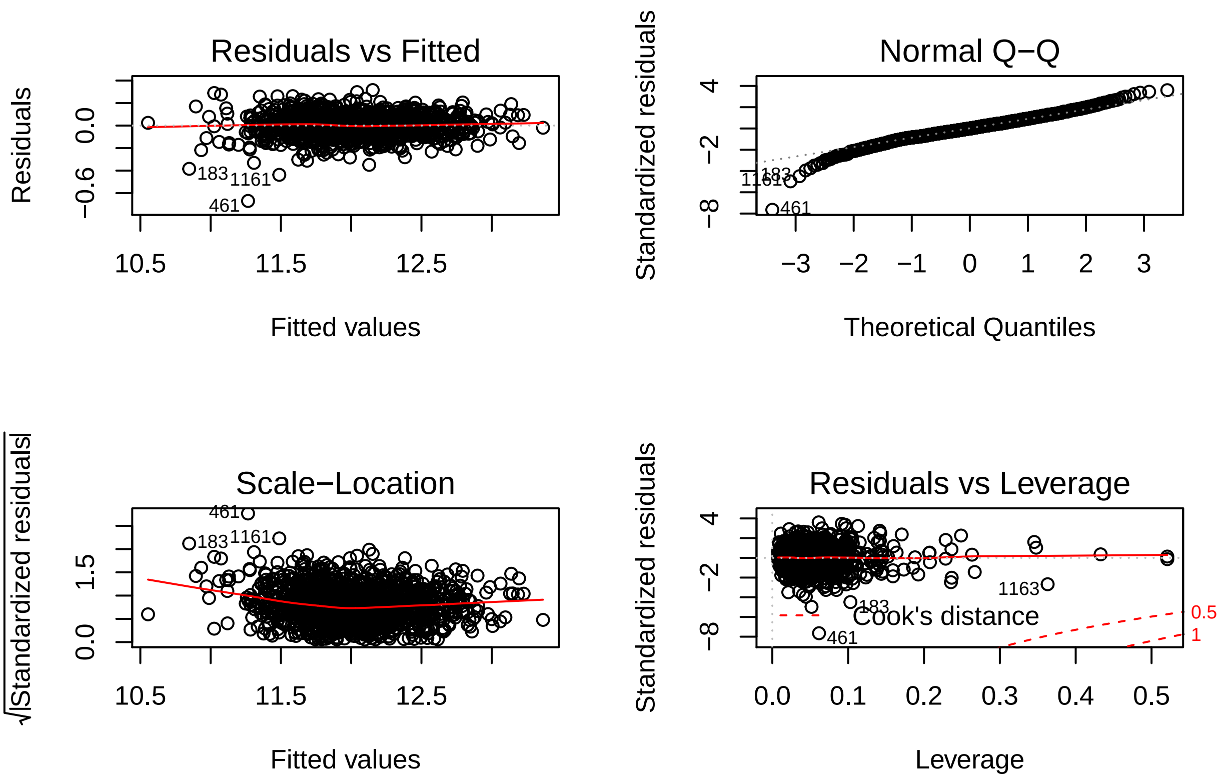
The 18 most important predictors are summarized in the following table.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	6.5102614	0.4057490	16.045046	0
log(Lot.Area)	0.0987836	0.0075415	13.098737	0
log(Total.Bsmt.SF)	0.1189555	0.0091090	13.059147	0
Overall.Cond	0.0336221	0.0027929	12.038585	0
Baths	0.0468993	0.0045396	10.331210	0

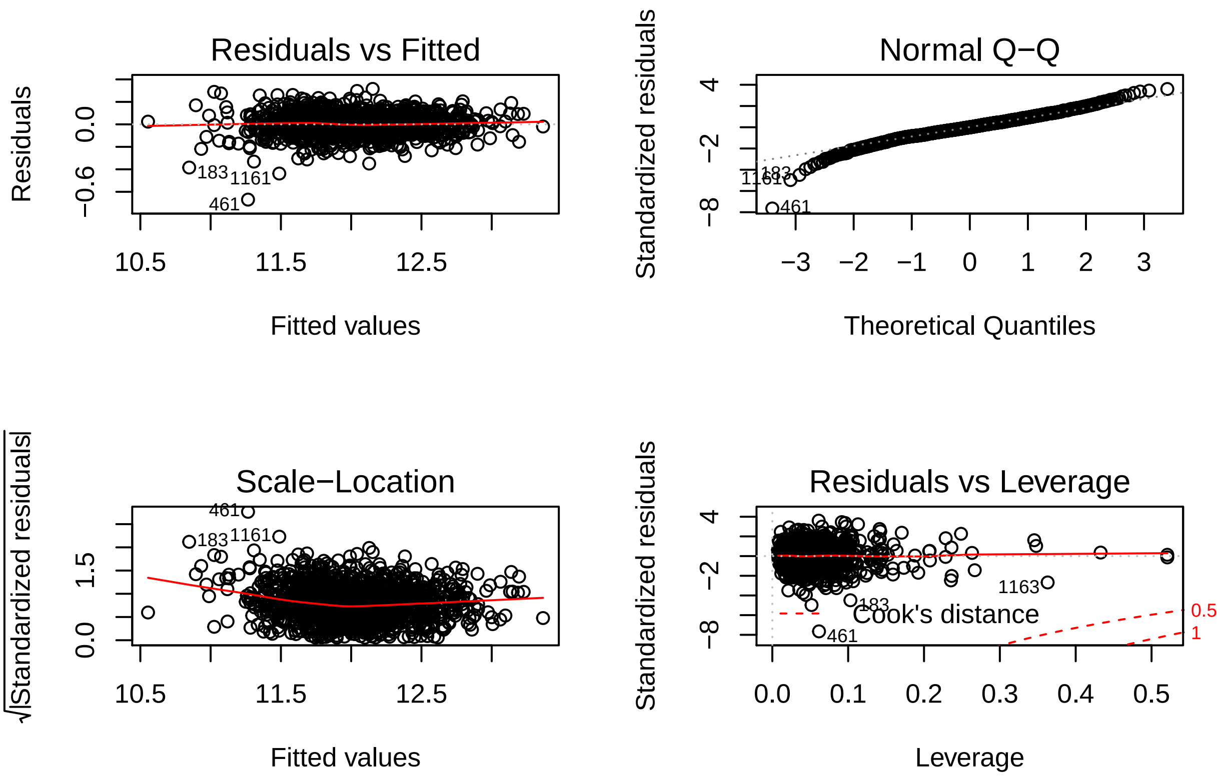
	Estimate	Std. Error	t value	Pr(>|t|)
log(TotalSq)	0.5012898	0.0581418	8.621853	0
Bsmt.QualTA	-0.8376454	0.1112764	-7.527611	0
Bsmt.QualGd	-0.8267350	0.1113854	-7.422292	0
Bsmt.QualFa	-0.8262069	0.1114845	-7.410957	0
Bsmt.QualEx	-0.7632015	0.1127644	-6.768105	0
Condition.1Feedr	-0.0745876	0.0112824	-6.610950	0
Kitchen.AbvGr	-0.0910776	0.0142064	-6.411031	0
FoundationPConc	0.0731980	0.0114720	6.380555	0
NeighborhoodGrnHill	0.4438689	0.0705238	6.293886	0
Fireplaces	0.0296062	0.0047109	6.284618	0
Kitchen.Qual	0.0332874	0.0055725	5.973465	0
Bsmt.QualPo	-0.8358514	0.1430638	-5.842509	0
FoundationCBlock	0.0589952	0.0104102	5.667074	0

As one would expect, and similarly to the conclusions from the simple model, we find that the overall condition of the house, lot area, and square footage are critical predictors of the house price. Interestingly, the square footage of the basement is an independent predictor, as is the number of bathrooms (Marc was particularly fascinated with the bathrooms).

Finally, here are the model analytics for the complete model.



4. Assessment of the final model (25 points)




```
##
## Call:
## lm(formula = log(price) ~ log(Lot.Area) + Condition.1 + Overall.Qual +
##     Baths + Neighborhood + Garage.Cars + log(Total.Bsmt.SF) +
##     log(TotalSq) + Overall.Cond + HouseAge + Foundation + Bsmt.Qual +
##     Bsmt.Exposure + Heating.QC + Central.Air + Bedroom.AbvGr +
##     Kitchen.AbvGr + Kitchen.Qual + Functional + Fireplaces +
##     Paved.Drive + Overall.Qual:Garage.Cars + Overall.Qual:log(TotalSq) +
##     Garage.Cars:log(TotalSq) + Overall.Qual:Garage.Cars:log(TotalSq),
##     data = data_train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.66963 -0.05131  0.00028  0.05415  0.31533
##
## Coefficients:
##                                        Estimate Std. Error t value
## (Intercept)                           6.5102614  0.4057490  16.045
## log(Lot.Area)                         0.0987836  0.0075415  13.099
## Condition.1Artery                    -0.0736717  0.0150310  -4.901
## Condition.1Feedr                     -0.0745876  0.0112824  -6.611
## Condition.1Park                      -0.0051531  0.0181031  -0.285
## Condition.1Rail                      -0.0456100  0.0148189  -3.078
## Overall.Qual                          0.2068733  0.0752569   2.749
## Baths                                 0.0468993  0.0045396  10.331
## NeighborhoodBlueste                   0.0203947  0.0416095   0.490
## NeighborhoodBrDale                   -0.0722602  0.0345132  -2.094
```

```
## NeighborhoodBrkSide                      -0.0217135  0.0304547  -0.713
## NeighborhoodClearCr                       0.0103092  0.0336539   0.306
## NeighborhoodCollgCr                      -0.0159676  0.0273774  -0.583
## NeighborhoodCrawfor                       0.0499783  0.0308951   1.618
## NeighborhoodEdwards                      -0.1032780  0.0291913  -3.538
## NeighborhoodGilbert                      -0.0368491  0.0285011  -1.293
## NeighborhoodGreens                        0.0574166  0.0441283   1.301
## NeighborhoodGrnHill                       0.4438689  0.0705238   6.294
## NeighborhoodIDOTRR                       -0.1470244  0.0322360  -4.561
## NeighborhoodLandmrk                      -0.0700365  0.0941311  -0.744
## NeighborhoodMeadowV                      -0.1441208  0.0380517  -3.787
## NeighborhoodMitchel                      -0.0346045  0.0293105  -1.181
## NeighborhoodNAmes                        -0.0586303  0.0283935  -2.065
## NeighborhoodNoRidge                       0.0250677  0.0307515   0.815
## NeighborhoodNPkVill                      -0.0218909  0.0381536  -0.574
## NeighborhoodNridgHt                       0.0317326  0.0289648   1.096
## NeighborhoodNWAmes                       -0.0465829  0.0295620  -1.576
## NeighborhoodOldTown                      -0.1206424  0.0292366  -4.126
## NeighborhoodSawyer                       -0.0277534  0.0297174  -0.934
## NeighborhoodSawyerW                      -0.0496333  0.0287674  -1.725
## NeighborhoodSomerst                       0.0576101  0.0274356   2.100
## NeighborhoodStoneBr                       0.0743440  0.0318815   2.332
## NeighborhoodSWISU                        -0.0658180  0.0339123  -1.941
## NeighborhoodTimber                       -0.0245197  0.0315509  -0.777
## NeighborhoodVeenker                      -0.0060078  0.0372772  -0.161
## Garage.Cars                               1.2032426  0.2341509   5.139
## log(Total.Bsmt.SF)                        0.1189555  0.0091090  13.059
## log(TotalSq)                              0.5012898  0.0581418   8.622
## Overall.Cond                              0.0336221  0.0027929  12.039
## HouseAge                                 -0.0009157  0.0001870  -4.896
## FoundationCBlock                          0.0589952  0.0104102   5.667
## FoundationPConc                           0.0731980  0.0114720   6.381
## FoundationSlab                            0.0718932  0.0289054   2.487
## FoundationStone                           0.0044701  0.0389376   0.115
## FoundationWood                            0.0392228  0.0550548   0.712
## Bsmt.QualEx                              -0.7632015  0.1127644  -6.768
## Bsmt.QualFa                              -0.8262069  0.1114845  -7.411
## Bsmt.QualGd                              -0.8267350  0.1113854  -7.422
## Bsmt.QualPo                              -0.8358514  0.1430638  -5.843
## Bsmt.QualTA                              -0.8376454  0.1112764  -7.528
## Bsmt.ExposureAv                           0.1575361  0.0910068   1.731
## Bsmt.ExposureGd                           0.1836363  0.0913334   2.011
## Bsmt.ExposureMn                           0.1122697  0.0912138   1.231
## Bsmt.ExposureNo                           0.1280950  0.0909443   1.408
## Heating.QC                                0.0093065  0.0033440   2.783
## Central.AirY                              0.0636338  0.0121613   5.232
## Bedroom.AbvGr                            -0.0147371  0.0042642  -3.456
## Kitchen.AbvGr                            -0.0910776  0.0142064  -6.411
## Kitchen.Qual                              0.0332874  0.0055725   5.973
## FunctionalMaj2                           -0.1605476  0.0514362  -3.121
## FunctionalMin1                            0.0184235  0.0343702   0.536
## FunctionalMin2                            0.0407226  0.0339893   1.198
## FunctionalMod                             0.0029487  0.0370700   0.080
## FunctionalTyp                             0.0888654  0.0310927   2.858
```

```
## Fireplaces                            0.0296062  0.0047109   6.285
## Paved.DriveP                          0.0171341  0.0181069   0.946
## Paved.DriveY                          0.0605392  0.0114220   5.300
## Overall.Qual:Garage.Cars             -0.1928486  0.0384268  -5.019
## Overall.Qual:log(TotalSq)            -0.0225185  0.0105342  -2.138
## Garage.Cars:log(TotalSq)             -0.1633273  0.0325288  -5.021
## Overall.Qual:Garage.Cars:log(TotalSq)  0.0269148  0.0052360   5.140
##                                      Pr(>|t|)
## (Intercept)                           < 2e-16 ***
## log(Lot.Area)                         < 2e-16 ***
## Condition.1Artery                    1.06e-06 ***
## Condition.1Feedr                     5.39e-11 ***
## Condition.1Park                      0.775950
## Condition.1Rail                      0.002125 **
## Overall.Qual                         0.006055 **
## Baths                                 < 2e-16 ***
## NeighborhoodBlueste                  0.624106
## NeighborhoodBrDale                   0.036463 *
## NeighborhoodBrkSide                  0.475977
## NeighborhoodClearCr                  0.759399
## NeighborhoodCollgCr                  0.559823
## NeighborhoodCrawfor                  0.105954
## NeighborhoodEdwards                  0.000416 ***
## NeighborhoodGilbert                  0.196256
## NeighborhoodGreens                   0.193425
## NeighborhoodGrnHill                  4.11e-10 ***
## NeighborhoodIDOTRR                   5.53e-06 ***
## NeighborhoodLandmrk                  0.456981
## NeighborhoodMeadowV                  0.000159 ***
## NeighborhoodMitchel                  0.237952
## NeighborhoodNAmes                    0.039111 *
## NeighborhoodNoRidge                  0.415112
## NeighborhoodNPkVill                  0.566222
## NeighborhoodNridgHt                  0.273459
## NeighborhoodNWAmes                   0.115302
## NeighborhoodOldTown                  3.90e-05 ***
## NeighborhoodSawyer                   0.350509
## NeighborhoodSawyerW                  0.084685 .
## NeighborhoodSomerst                  0.035919 *
## NeighborhoodStoneBr                  0.019846 *
## NeighborhoodSWISU                    0.052476 .
## NeighborhoodTimber                   0.437201
## NeighborhoodVeenker                  0.871985
## Garage.Cars                          3.15e-07 ***
## log(Total.Bsmt.SF)                    < 2e-16 ***
## log(TotalSq)                          < 2e-16 ***
## Overall.Cond                          < 2e-16 ***
## HouseAge                             1.09e-06 ***
## FoundationCBlock                     1.76e-08 ***
## FoundationPConc                      2.38e-10 ***
## FoundationSlab                       0.012990 *
## FoundationStone                      0.908619
## FoundationWood                       0.476314
## Bsmt.QualEx                          1.90e-11 ***
```

```
## Bsmt.QualFa                             2.14e-13 ***
## Bsmt.QualGd                             1.97e-13 ***
## Bsmt.QualPo                             6.36e-09 ***
## Bsmt.QualTA                             9.13e-14 ***
## Bsmt.ExposureAv                         0.083662 .
## Bsmt.ExposureGd                         0.044555 *
## Bsmt.ExposureMn                         0.218586
## Bsmt.ExposureNo                         0.159202
## Heating.QC                              0.005457 **
## Central.AirY                            1.92e-07 ***
## Bedroom.AbvGr                           0.000564 ***
## Kitchen.AbvGr                           1.96e-10 ***
## Kitchen.Qual                            2.93e-09 ***
## FunctionalMaj2                          0.001837 **
## FunctionalMin1                          0.592021
## FunctionalMin2                          0.231077
## FunctionalMod                           0.936611
## FunctionalTyp                           0.004324 **
## Fireplaces                              4.36e-10 ***
## Paved.DriveP                            0.344168
## Paved.DriveY                            1.34e-07 ***
## Overall.Qual:Garage.Cars                5.86e-07 ***
## Overall.Qual:log(TotalSq)               0.032715 *
## Garage.Cars:log(TotalSq)                5.79e-07 ***
## Overall.Qual:Garage.Cars:log(TotalSq) 3.12e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09038 on 1422 degrees of freedom
## Multiple R-squared:  0.9437, Adjusted R-squared:  0.941
## F-statistic: 340.7 on 70 and 1422 DF,  p-value: < 2.2e-16
```

- Residual:

  Pardoe's paper "Modeling home prices using realtor data" suggests an issue of hetereoscedasticity with variation increasing with sale price. While it caused us concern in the beginning, it was not an issue after we had logged the response variable. The residual plot shows that the variance is relatively constant and no obvious patterns exist, which suggests that the logged response helped account for this non-constant variance. However, there are some outliers in the bottom-left of the plot where properties are overvalued. In addition, the variance in the left part of the plot is slightly higher in predicted prices. We do not throw away outliers in the model.

- Model Evaluation:

  From the summary table, our model has a mutiple R-squared of 0.9437, which means that the final model explains about 94 percent of variation in the data. In addition, the residual standard error of 0.09 indicates that the fit on training data is decent. The diagnostic plots suggest that our model actually did a better job in predicting price for those properties in middle and high price ranges than lower price range. The residual plot has higher variance in the lower price range. In addition, the normal QQ plot generally follows a straight line, but with a heavier left tail. In the residual vs leverage plot, it can be observed that there are about 5 high leverage point, but they are not influential because no points have cook's distance greater than 0.5. In conclusion, our final model did a good job in predicting the prices, especially for the properties in middle and high price ranges.

```
##        bias  max.dev mean.dev    RMES Coverage
## 1 -156.1364 64352.67 11229.45 14983.7    0.966
```

- RMSE discussion and Model Testing:

  The RMSE for our final model is 14983.7. After comparing with other groups in the leaderboard, we find that the bias for our model is actually the lowest. However, due to the variance and bias trade-off, our model has a higher deviance or variance for prediction, possibly resulting from outliers in the dataset. Another possible issue would be extrapolation, which affects the prediction accuracy for some properties with extreme prices. The coverage of prediction of 0.966 indicates that our model satisfactorily captures true prices within the prediction interval.

PID pr	ice pri	ce_to_pred_ratio Tot	alSq Ove	rall.Qual Neig	hborhood
535382020	160000	1.334928	2414	5	OldTown
905452050	113000	1.253250	672	4	Edwards
535454070	166000	1.381293	1385	5	NAmes
903400220	214500	1.251595	2134	6	BrkSide
910206010	68104	1.351226	640	2	IDOTRR
916226030	241500	1.217801	1501	7	Timber
910203020	120500	1.208631	778	5	IDOTRR
903429110	179900	1.228883	1944	7	OldTown
905427010	235000	1.243933	2009	6	Edwards
905376090	216000	1.424357	1325	5	Edwards

PID pr	ice pri	ce_to_pred_ratio Tot	alSq Ove	rall.Qual Neig	hborhood
528102010	315000	0.8440589	1980	9	NridgHt
535150070	104900	0.8279967	1738	4	NAmes
902105130	64000	0.7702524	672	5	OldTown
534451130	79900	0.7688397	747	4	BrkSide
909101010	110000	0.7873724	1196	6	Edwards
532378110	127000	0.8488299	1040	5	Sawyer
905200160	80000	0.7450666	1006	5	Sawyer
532351140	112000	0.8317835	1902	6	Sawyer
527182020	130000	0.7556536	1204	8	StoneBr
534479120	105000	0.8349797	1376	5	NAmes

- Model result:

  The two tables show the top 10 most under- and over-priced properties based on our final model. The real-to-prediction ratio over 1 suggests over-priced and less than 1 under-priced. We also include some other features to compare the under and over-priced properties. One of our teammates comments that, were he to be a property investor in Ames, Iowa, he would keep an eye on the Sawyer neighborhood for buying opportunities, given that there were three here in the undervalued top 10, while considering selling properties that were in the Edwards neighborhood, as there were three here in the overvalued top 10.

5. Conclusion (10 points)

In essence, the overall house quality, neighborhood, and total square footage were the most important predictors in our model. While this makes sense intuitively, confirming these predictors' relative importance in a quantitative analysis is reassuring.

Interestingly, the number of bathrooms is an independent predictor of house price. Who would have guessed it? Not us!

In reflecting upon the work we have done and our results, we have learned several things about the data analysis process, as well as the pricing of houses.

One of the key aspects that was apparent to us was how important our data-cleaning was to creating a successful model. In the early stages we had fit a model using the data more or less in its original form, with few modifications. After our meticulous data-cleaning session, the improvements were dramatic, as our bias dropped drastically and RMSE considerably.

Another observation we take away from our project is the realization that OLS is often very useful, and that more advanced methods are not always optimal. Over the course of the project we attempted running Ridge, Lasso, Blasso, Trees, Random Forests, and Boosting, with linear regression of the same variables ultimately outperforming them all.

Another thing we took away was the usefulness of tree models, not only in prediction, but also during model selection. We found boosting to be particularly useful in finalizing the variables we wanted to keep in our model through the examining the relative importance plot, and found a decision tree to be useful in indicating to us important interactions to consider in constructing our complex model.

Part IV

Create predictions for the validation data from your final model and write out to a file `prediction-validation.Rdata` This should have the same format as the models in Part I and II.

Please see prediction-validation for predictions.

```
##          fit       lwr       upr       PID
## 1  101107.67  84294.37 121274.54 527451060
## 2  130025.33 108593.31 155687.19 531452020
## 3  153002.06 127946.03 182964.89 907126010
## 4  104755.81  87447.58 125489.81 535379060
## 5  138568.77 115939.12 165615.41 535303110
## 6  137038.09 114495.38 164019.17 907227290
## 7   95104.43  79324.55 114023.39 903458170
## 8  375836.19 313246.44 450931.98 528477050
## 9  235855.61 197259.22 282003.89 907250020
## 10 189588.45 158577.03 226664.49 527166030
## 11 198947.84 165714.03 238846.66 527375100
## 12 122645.86 102283.02 147062.62 903429010
## 13 184380.84 154102.80 220607.90 527163070
## 14 125678.44 105130.17 150242.99 535478010
## 15 119447.51  99165.27 143878.09 923225200
## 16 175665.02 146560.66 210548.99 903400180
## 17 115086.78  95978.08 137999.91 902128100
## 18 175261.47 146402.39 209809.30 527325240
## 19 201729.05 168404.50 241647.99 909282110
## 20 224234.03 186918.47 268999.10 907131090
## 21 281875.00 235320.13 337640.11 528142130
## 22 165510.57 137796.17 198799.06 909177120
## 23 206048.89 172234.34 246502.21 907175060
## 24 193750.86 161973.12 231763.13 527359070
## 25 226490.17 189258.94 271045.57 906378110
## 26 155225.00 129079.58 186666.23 531475220
## 27 164163.19 137393.43 196148.78 534400060
## 28  89769.74  74575.78 108059.29 923228260
## 29  87221.25  72454.69 104997.31 904302020
## 30 318986.32 265590.37 383117.33 528429090
```

```
## 31  132111.79 110042.48 158607.15 905475150
## 32  179864.77 150199.64 215388.90 527302070
## 33  117368.11  97968.31 140609.49 903231190
## 34  152898.34 127927.95 182742.72 534428100
## 35  180613.76 151032.23 215989.18 907131060
## 36  288589.73 240910.86 345704.77 906340120
## 37  149512.57 124268.98 179884.05 909252010
## 38  138387.63 115643.93 165604.35 535175070
## 39  105451.62  87829.57 126609.34 532376170
## 40  154465.56 128755.43 185309.54 905108090
## 41  114143.01  95379.82 136597.31 902406030
## 42  178482.84 149229.13 213471.21 905103040
## 43  123248.98 103095.96 147341.48 535304100
## 44  231346.80 192013.48 278737.42 533250050
## 45  358877.36 298959.75 430803.67 916382100
## 46  180587.22 150923.19 216081.73 527107090
## 47  319362.30 266607.99 382555.23 528344040
## 48  128671.25 107548.64 153942.36 534201130
## 49  151502.81 126239.70 181821.58 903450060
## 50  185295.35 153867.86 223141.91 902105020
## 51  139925.20 116387.23 168223.45 903204030
## 52  140127.02 117169.51 167582.68 907290170
## 53  131678.48 110049.44 157558.47 907200170
## 54  199931.40 167015.72 239334.15 528253010
## 55  128144.39 107067.76 153370.03 905108190
## 56  124532.77 103864.20 149314.30 903481120
## 57  203018.19 169215.48 243573.37 916455010
## 58  126289.01 105302.65 151457.84 902105050
## 59  143226.67 119720.19 171348.55 905452160
## 60  164654.70 137659.99 196942.99 534127270
## 61  117707.94  98415.31 140782.56 905476225
## 62  138088.02 115285.38 165400.86 923277040
## 63  159056.32 132454.09 191001.38 909275020
## 64  133127.37 111224.61 159343.30 903233180
## 65  139584.10 116415.60 167363.49 905478140
## 66  199748.12 167122.64 238742.72 907412020
## 67  182135.49 152148.59 218032.49 527354200
## 68  366561.43 303314.67 442996.31 902400110
## 69  144567.54 120532.62 173395.17 905200490
## 70   86843.69  72179.46 104487.17 923225300
## 71  157298.78 131448.74 188232.37 923201020
## 72  124340.88 103662.34 149144.37 532351050
## 73  149142.03 124352.30 178873.62 531363060
## 74  226876.93 189012.32 272326.92 534128010
## 75  188524.28 157290.39 225960.44 909250150
## 76  180074.25 150608.33 215305.05 528290180
## 77  142468.85 118903.16 170705.09 531451150
## 78  252546.85 210731.82 302659.12 909282020
## 79   79460.09  66012.29  95647.44 903481100
## 80  150962.95 125906.64 181005.64 905107280
## 81  169651.02 141843.73 202909.70 527162090
## 82  271479.67 226673.69 325142.32 528478030
## 83  123623.28 103274.58 147981.38 904100040
## 84  198874.26 166375.83 237720.67 907260040
```

```
## 85   198354.32 165100.48 238306.01 916226090
## 86   109060.68  91126.18 130524.86 905104210
## 87   218813.85 182592.12 262221.06 906201120
## 88   220711.87 184558.95 263946.72 907418010
## 89   177535.91 147262.65 214032.54 533254100
## 90   173956.84 145385.98 208142.36 533217060
## 91   112301.02  93437.12 134973.34 903475060
## 92   268772.87 224355.20 321984.31 528344060
## 93   141582.02 118248.72 169519.53 531452010
## 94   180482.83 150163.51 216923.89 528228440
## 95   187862.04 156927.39 224894.74 903236130
## 96   150038.78 125341.95 179601.75 907202010
## 97   196199.06 163578.97 235324.08 527145080
## 98   156286.33 130382.54 187336.55 909277100
## 99   302504.01 252642.07 362206.82 528382020
## 100 108742.28  90829.49 130187.73 908102290
## 101 149647.81 125103.62 179007.34 535453020
## 102 146533.16 122396.67 175429.33 905108170
## 103 154858.27 128212.02 187042.39 527226010
## 104 146000.82 121865.90 174915.53 909281020
## 105 178363.76 148404.55 214370.98 528228465
## 106 284885.73 238029.63 340965.45 528431120
## 107 127516.82 106439.65 152767.70 910202070
## 108 149671.32 125081.60 179095.12 902201140
## 109 414410.96 345505.02 497059.19 528327010
## 110 158955.86 133003.90 189971.60 535153150
## 111 195792.02 163568.85 234363.18 914476330
## 112 203975.82 170038.57 244686.45 909428180
## 113 133529.42 111337.91 160144.08 903454010
## 114 202420.01 169209.21 242149.12 528490080
## 115  83280.99  69259.49 100141.12 903456060
## 116 251965.20 210734.05 301263.43 907251090
## 117  81755.42  68181.47  98031.76 902125160
## 118  85326.17  70874.14 102725.14 923228220
## 119  82702.66  68798.99  99416.15 903232170
## 120 116306.99  96790.61 139758.55 910201050
## 121 103670.96  86456.82 124312.55 535353130
## 122 329055.98 274934.79 393830.99 528142140
## 123 143649.67 119937.38 172050.00 924100050
## 124 120413.32 100713.95 143965.84 534276040
## 125 222968.49 186457.77 266628.44 907265030
## 126 223596.59 186904.77 267491.50 906394040
## 127 122785.45 102298.40 147375.40 908130020
## 128 223227.21 186686.36 266920.33 906380050
## 129 175016.73 146318.94 209343.07 526352090
## 130 140527.20 117346.84 168286.55 903227090
## 131 166391.13 138634.51 199705.03 924100020
## 132  96680.14  80400.86 116255.59 902427140
## 133 217697.56 182006.02 260388.24 528480130
## 134 209896.94 175315.21 251300.07 531369010
## 135 314491.17 262662.56 376546.61 528315110
## 136 358726.01 299059.12 430297.36 528150110
## 137 198409.60 164869.25 238773.27 909251090
## 138 295759.76 247042.52 354084.12 528180070
```

```
## 139 191039.60 159215.01 229225.43 526301100
## 140 144846.24 121095.30 173255.54 908152070
## 141 164776.22 136942.22 198267.57 909275250
## 142 228803.20 190837.76 274321.50 909281080
## 143 111553.90  92982.68 133834.32 910202100
## 144 263437.78 219337.19 316405.37 911370520
## 145 106503.26  88791.91 127747.49 527450030
## 146 145178.01 121143.86 173980.38 535451250
## 147 219461.71 183568.46 262373.18 907285080
## 148 138779.46 115314.99 167018.52 909176180
## 149  86716.91  72434.74 103815.15 535375010
## 150 220000.75 183911.21 263172.25 528235160
## 151 139420.91 116319.19 167110.78 909101140
## 152 165000.43 137575.76 197891.99 905226050
## 153  81464.80  67647.51  98104.32 923228270
## 154 126404.90 104864.25 152370.31 923225040
## 155 335823.11 279810.09 403048.93 916475040
## 156 135834.49 113245.43 162929.40 534250010
## 157 238544.70 199077.40 285836.44 909475050
## 158 131000.78 109385.55 156887.31 903226150
## 159 168397.32 140844.45 201340.26 528240040
## 160 144623.82 120761.50 173201.31 534225110
## 161 217918.41 181618.14 261474.07 906475100
## 162 130716.50 109131.12 156571.33 923230180
## 163 304757.58 254491.97 364951.32 528326030
## 164 136776.97 114096.88 163965.39 903227140
## 165 214768.24 179607.54 256812.14 907187060
## 166 200912.83 167956.90 240335.25 528480110
## 167 123529.59 102780.57 148467.35 908225290
## 168 213375.10 177965.40 255830.26 924100040
## 169 122124.59 102056.18 146139.27 923225490
## 170 108689.96  90208.63 130957.61 911226010
## 171 110911.07  92387.53 133148.53 904351200
## 172 133107.86 111141.06 159416.35 903231180
## 173 140968.58 117597.34 168984.62 903204040
## 174  90922.29  75120.18 110048.49 903232030
## 175 237377.48 198345.05 284091.11 907295040
## 176 149081.76 124367.74 178706.87 909252150
## 177 263590.21 220117.82 315648.21 533128030
## 178 144685.82 120843.06 173232.83 914475010
## 179 103784.22  86454.55 124587.60 905101100
## 180 149392.79 124870.26 178731.16 533223100
## 181 129668.43 107998.66 155686.19 902103120
## 182 113355.52  94047.97 136626.81 923225080
## 183 170702.72 142774.88 204093.47 527163020
## 184 131020.41 109564.77 156677.63 532377140
## 185 249160.28 206997.37 299911.26 911370510
## 186 131546.85 109938.29 157402.61 907200340
## 187  56489.91  46859.84  68099.05 910203100
## 188 129556.43 108084.36 155294.15 902206090
## 189 121894.54 101930.83 145768.25 905451300
## 190 120619.78 100624.54 144588.31 526302110
## 191 175471.15 146760.61 209798.28 535154060
## 192 129646.51 108367.29 155104.16 535456110
```

```
## 193 268071.51 223917.90 320931.62 533215020
## 194 130075.39 108585.09 155818.90 902105060
## 195 167152.79 139716.97 199976.09 907420110
## 196 183908.67 153693.44 220064.05 528240060
## 197 112802.12  94099.90 135221.38 905226030
## 198 139433.74 116322.63 167136.58 903401050
## 199 277930.92 231985.22 332976.36 906380100
## 200 120166.61 100268.63 144013.27 902103150
## 201 234268.49 195165.30 281206.37 534128210
## 202 178239.57 148985.67 213237.59 531452080
## 203 141450.87 118169.67 169318.82 914452190
## 204 222918.27 186376.91 266623.99 906223090
## 205 280949.79 233483.99 338065.08 533250160
## 206 237042.59 198134.65 283590.94 528482130
## 207 115048.51  96275.38 137482.29 527404100
## 208 115143.02  96267.80 137719.10 532378120
## 209 221232.68 184922.22 264672.88 906223140
## 210 312826.62 260417.04 375783.77 527146135
## 211 169177.87 141222.21 202667.50 534251280
## 212 137396.19 114236.34 165251.38 534453150
## 213 138914.18 115993.20 166364.50 908127060
## 214  72262.49  60268.18  86643.86 910203230
## 215 120750.50 100427.98 145185.48 903454100
## 216 156127.15 129859.10 187708.73 905201120
## 217 269915.89 225381.52 323250.04 528106120
## 218 118722.97  98956.02 142438.46 527451380
## 219 179712.60 149655.99 215805.71 906475050
## 220 169903.63 142130.67 203103.55 526351100
## 221 189154.77 157887.59 226613.93 528174030
## 222 141289.39 118231.62 168843.95 535152250
## 223 216656.37 181121.35 259163.16 528490070
## 224 208420.61 174008.44 249638.18 534152070
## 225 151294.16 124381.84 184029.46 923225260
## 226 136509.59 114190.78 163190.66 527402150
## 227 213329.11 178072.36 255566.39 528181070
## 228 233846.08 195208.22 280131.59 907275010
## 229 202239.80 167133.72 244719.84 533253030
## 230 231233.59 193175.42 276789.73 534177180
## 231  74529.83  61788.57  89898.42 923226180
## 232 169074.42 140391.27 203617.79 533251120
## 233 195555.08 163513.64 233875.22 534403290
## 234 144611.18 120965.85 172878.50 535402070
## 235 202531.43 169416.08 242119.77 907194110
## 236 215406.66 179790.25 258078.67 909452050
## 237 135000.17 112096.53 162583.50 527452060
## 238 241913.02 202074.77 289605.22 907254020
## 239 114107.41  95304.38 136620.18 905403060
## 240 176096.90 146846.78 211173.30 534127130
## 241 114948.54  96044.38 137573.57 902427150
## 242  94276.93  78772.47 112833.07 527403120
## 243 266007.83 222149.37 318525.16 528354050
## 244 239011.16 199924.73 285739.20 907262060
## 245 113195.79  93994.39 136319.70 903476090
## 246 231282.79 192935.90 277251.29 527366030
```

```
## 247  99905.50  83262.33 119875.45 902108060
## 248 138064.02 115465.52 165085.40 905106210
## 249 123785.21 103511.91 148029.11 902201110
## 250 189800.54 158551.85 227207.96 905200280
## 251 122607.16 102528.93 146617.30 535327160
## 252 117443.16  97922.96 140854.57 908102330
## 253 122575.27 102278.03 146900.54 902206130
## 254 137835.67 114951.37 165275.73 535477060
## 255 126248.63 105591.80 150946.53 535302080
## 256 245590.07 204989.86 294231.53 528439060
## 257 177829.47 148641.83 212748.45 533223020
## 258 133964.15 111941.15 160319.89 532377130
## 259 184770.83 153679.68 222152.08 905201110
## 260 126470.05 105568.46 151509.97 903231080
## 261 300949.59 251354.47 360330.39 528482090
## 262 125981.89 105164.06 150920.72 905104240
## 263 219021.05 182518.24 262824.26 527184110
## 264 109965.90  91983.77 131463.41 535452140
## 265 289025.98 240992.04 346633.93 905427030
## 266 140898.36 117785.52 168546.59 535456070
## 267 103010.05  86073.18 123279.64 527425090
## 268 183396.77 153401.93 219256.54 528294010
## 269 139172.18 116331.97 166496.76 527357110
## 270 249594.42 208595.64 298651.38 907260010
## 271 257031.77 214701.50 307707.83 533215070
## 272 141185.08 117919.89 169040.41 903234220
## 273 110370.83  91930.47 132510.14 902301060
## 274 112943.84  94260.94 135329.77 923275140
## 275 135023.01 112515.66 162032.67 531375090
## 276  95961.10  79861.72 115305.97 903204095
## 277 232582.00 193353.86 279768.84 909428240
## 278 294789.56 246258.92 352884.21 533130160
## 279 200467.75 166727.45 241036.01 528228325
## 280 135024.32 112820.95 161597.34 907202080
## 281 131393.24 109961.87 157001.54 527402220
## 282 351307.30 293509.94 420486.00 528142040
## 283 146944.83 122529.06 176225.82 535477130
## 284 211911.43 177063.72 253617.48 534151090
## 285 195781.70 160909.39 238211.55 902329030
## 286 206274.78 172322.49 246916.62 528186170
## 287 173987.20 145416.17 208171.79 531363010
## 288 186931.17 156358.44 223481.78 906380190
## 289 322670.81 268028.17 388453.41 909428280
## 290 238801.85 199390.84 286002.71 528358010
## 291 255835.97 213563.52 306475.76 528172080
## 292 172787.10 144189.21 207056.97 906201022
## 293 167656.22 140121.54 200601.63 533210060
## 294 133960.34 111317.01 161209.62 905227050
## 295 149845.75 125370.18 179099.60 534429130
## 296 181324.39 151692.16 216745.11 535402220
## 297 311819.50 259979.65 373996.21 921201060
## 298 108682.14  90606.84 130363.31 527450150
## 299  76086.93  63236.37  91548.91 908250030
## 300 120387.83 100166.41 144691.53 923250210
```

```
## 301 138225.85 115176.80 165887.46 902110080
## 302 151502.97 126693.41 181170.83 907135020
## 303 318696.09 266233.78 381496.28 528344020
## 304 259619.02 216823.54 310861.24 528315060
## 305 182531.43 152668.74 218235.41 907414080
## 306 299896.46 248834.98 361435.86 533350120
## 307 302920.44 252670.33 363164.11 916475100
## 308 197496.06 165202.77 236101.94 528240050
## 309  91519.80  76457.72 109549.08 902111010
## 310 215289.20 179983.30 257520.78 527368020
## 311 241771.63 201539.94 290034.42 528429050
## 312 110453.37  92130.73 132419.96 535478110
## 313 167260.76 139627.39 200362.99 526302120
## 314 232128.39 194057.94 277667.52 535125070
## 315 104580.22  87314.98 125259.41 902301150
## 316  95107.53  79211.94 114192.90 531477050
## 317 109268.36  91091.08 131072.94 902136110
## 318 278394.31 232266.47 333683.08 533110130
## 319 217718.29 181676.57 260910.10 534152050
## 320  65494.39  54517.46  78681.49 902456015
## 321 112907.36  93909.80 135748.04 903425420
## 322 190523.96 158394.06 229171.34 906476030
## 323 155899.41 130208.48 186659.32 923227030
## 324 180108.86 150645.42 215334.79 906394050
## 325 314685.75 262322.35 377501.65 527252090
## 326 172519.64 143911.76 206814.42 924100060
## 327 150192.50 125146.96 180250.37 902332030
## 328 319028.14 266276.96 382229.66 906340130
## 329 184626.31 154314.67 220891.99 533213020
## 330 163563.26 136737.04 195652.48 534401130
## 331 195271.88 163233.22 233598.94 527327050
## 332 199680.05 166860.50 238954.83 534126060
## 333 137161.11 114501.01 164305.71 905102060
## 334 140840.69 117756.60 168450.01 534430080
## 335 220511.71 184370.72 263737.19 907196050
## 336 210473.47 175819.11 251958.28 528138060
## 337 172703.22 144332.26 206650.96 531385070
## 338 171000.07 141839.00 206156.45 527226040
## 339 136656.20 113475.23 164572.62 527454120
## 340 107651.67  89902.32 128905.27 902109010
## 341 119289.11  99657.02 142788.66 903236200
## 342 132820.84 111104.51 158781.81 534431030
## 343 236490.02 197547.31 283109.55 906402060
## 344 132357.40 110438.24 158626.94 908186050
## 345 117436.70  98059.96 140642.30 535302140
## 346 133402.20 111567.78 159509.74 532378050
## 347 326869.74 271910.67 392937.24 905301050
## 348 166902.99 139274.80 200011.84 535300120
## 349 124177.74 103325.19 149238.64 908127070
## 350 200929.05 167940.71 240397.23 531451020
## 351 127437.66 106160.05 152979.93 909176150
## 352 209400.96 174786.92 250869.80 528235130
## 353 237413.18 197132.85 285924.03 533250130
## 354 118584.26  99020.08 142013.89 531478010
```

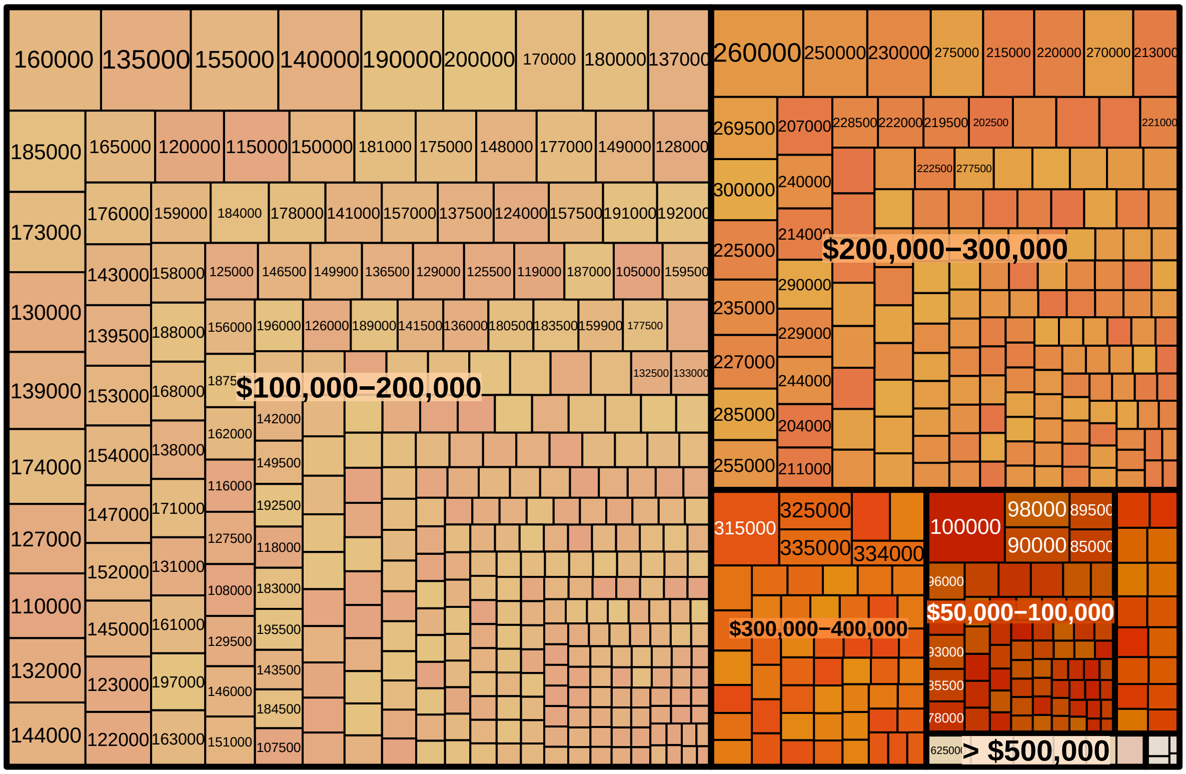
```
## 355 130518.46 109064.80 156192.17 907202100
## 356 179791.15 150030.35 215455.46 531452260
## 357 152790.53 127821.30 182637.38 534202170
## 358 176492.51 146936.37 211993.85 905225090
## 359 141577.32 117764.86 170204.74 532479050
## 360 230723.67 192635.94 276342.06 534151175
## 361 198660.23 165889.61 237904.52 528174010
## 362 116526.95  97498.12 139269.65 535404080
## 363 372990.07 311037.34 447282.59 528366050
## 364 209583.80 174795.92 251295.17 527302110
## 365 131787.46 109895.88 158039.90 905378040
## 366 153624.52 128082.30 184260.37 914452060
## 367 210263.85 175168.45 252390.70 905352180
## 368 227257.58 189637.83 272340.22 914474070
## 369 147827.50 123548.94 176877.04 534176250
## 370 200280.74 166923.87 240303.41 903456110
## 371 119562.15  99946.87 143027.06 905106170
## 372 138564.32 115877.05 165693.48 907135260
## 373 107382.35  89430.64 128937.58 904100190
## 374 428277.23 357002.07 513782.40 528340030
## 375 152944.93 127396.31 183617.19 903401070
## 376  68804.48  56718.41  83465.95 902402260
## 377 135162.24 112847.91 161888.97 907227060
## 378 230025.95 191654.50 276079.80 527182190
## 379 128380.68 107152.64 153814.23 534451110
## 380 115214.69  96166.11 138036.42 911175360
## 381 228973.42 191437.40 273869.31 907275090
## 382 147938.97 123055.96 177853.54 904302030
## 383 171384.58 142767.02 205738.51 923276250
## 384 219069.47 183010.18 262233.68 527354040
## 385 423891.37 353758.03 507928.80 528118040
## 386 208173.92 170175.86 254656.44 902326030
## 387  71947.61  59968.75  86319.27 909101060
## 388 194551.16 162776.62 232528.19 907192030
## 389  81863.23  68022.46  98520.24 923226290
## 390 161231.86 134530.49 193232.87 535351050
## 391 108385.38  89986.85 130545.63 911103060
## 392 149364.46 124465.86 179243.88 531375050
## 393 118454.94  98855.61 141940.08 534201300
## 394 214871.37 179749.72 256855.52 907405020
## 395 144595.79 120852.93 173003.20 907126050
## 396 275737.17 230263.99 330190.53 907280100
## 397 187964.86 156974.53 225073.37 528188040
## 398 164648.33 136721.01 198280.22 527453010
## 399 100872.92  84176.57 120880.98 902105010
## 400 186511.51 155926.84 223095.28 533223030
## 401 326664.17 272698.62 391309.20 528363020
## 402 219436.17 183459.22 262468.31 907180130
## 403 240764.80 201279.32 287996.24 528240080
## 404 141351.71 117702.66 169752.37 909176140
## 405 207486.96 173220.72 248531.69 528181050
## 406 173065.19 144579.28 207163.57 914476010
## 407 153703.89 128480.02 183879.85 535426350
## 408 157073.99 131078.81 188224.45 909254150
```

```
## 409 123005.30 102576.02 147503.32 527276150
## 410 145106.46 120549.66 174665.64 527455080
## 411 174608.86 145584.41 209419.77 527358090
## 412  99464.82  83051.21 119122.29 903235020
## 413 135456.48 113021.15 162345.35 532376110
```

Appendix

One team member was really into making graphs that were subsequently rejected by the rest of the team. As a token of appreciation to the energetic team member, we include their visualizations as an appendix.

Effect of Total Square Footage on Price Range



Price by Overall Quality