521 - Final Project*

Thomas Fleming, Eden Huang, Blaire Li, Marc Ryser

May 1, 2017

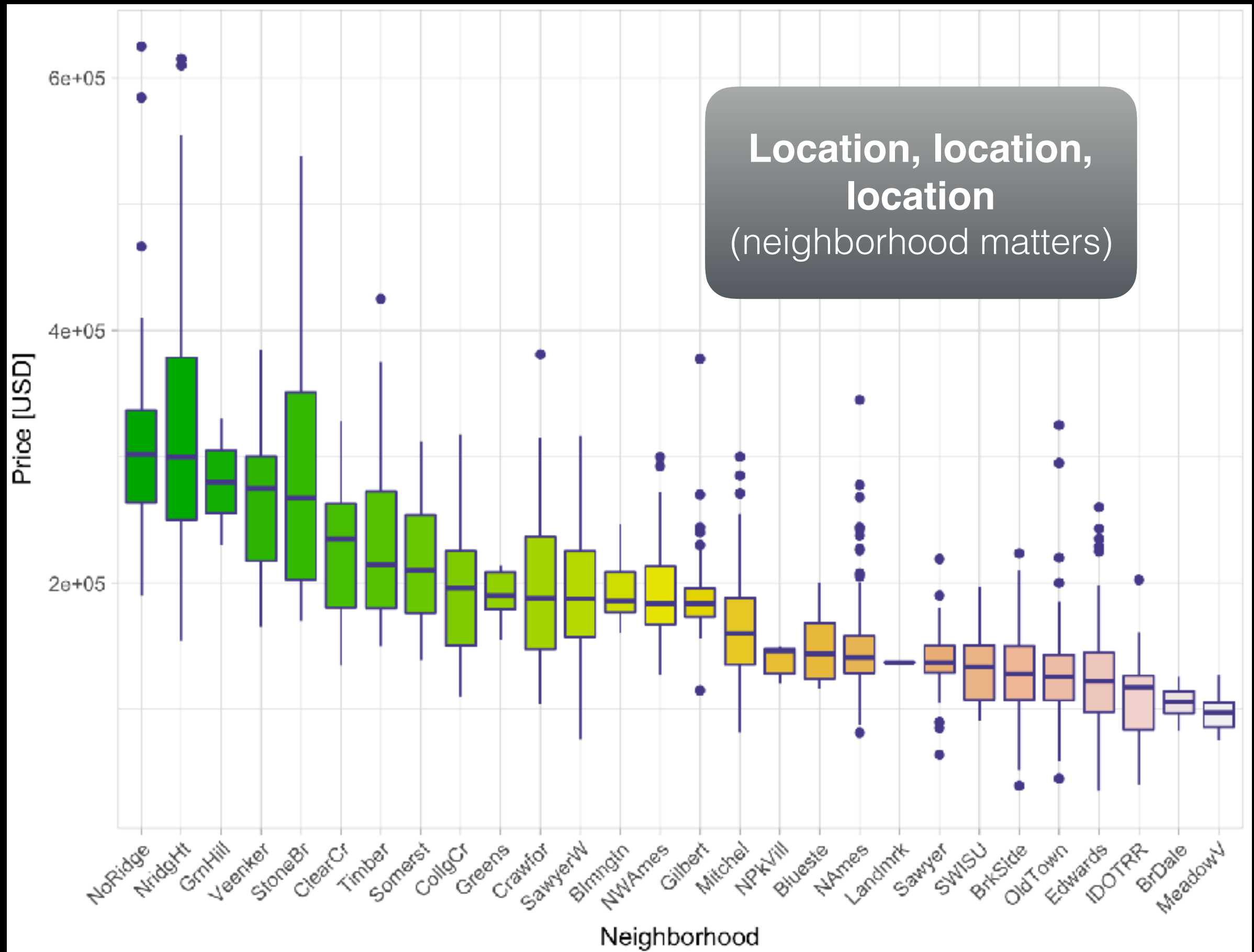*We'll make sure your dream home comes at the right price
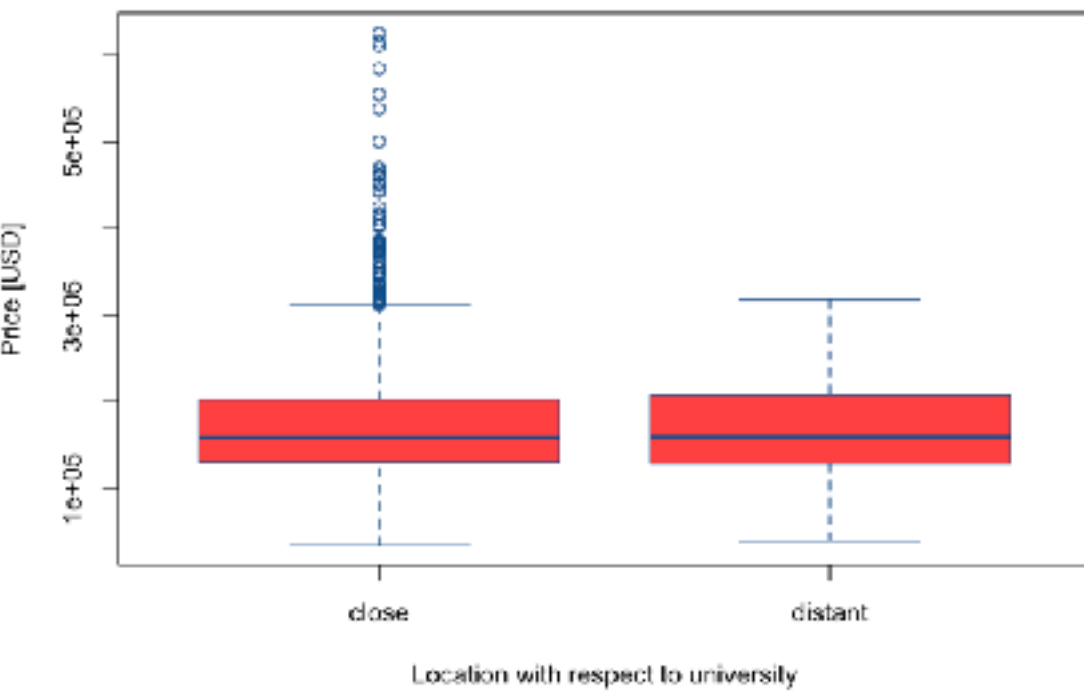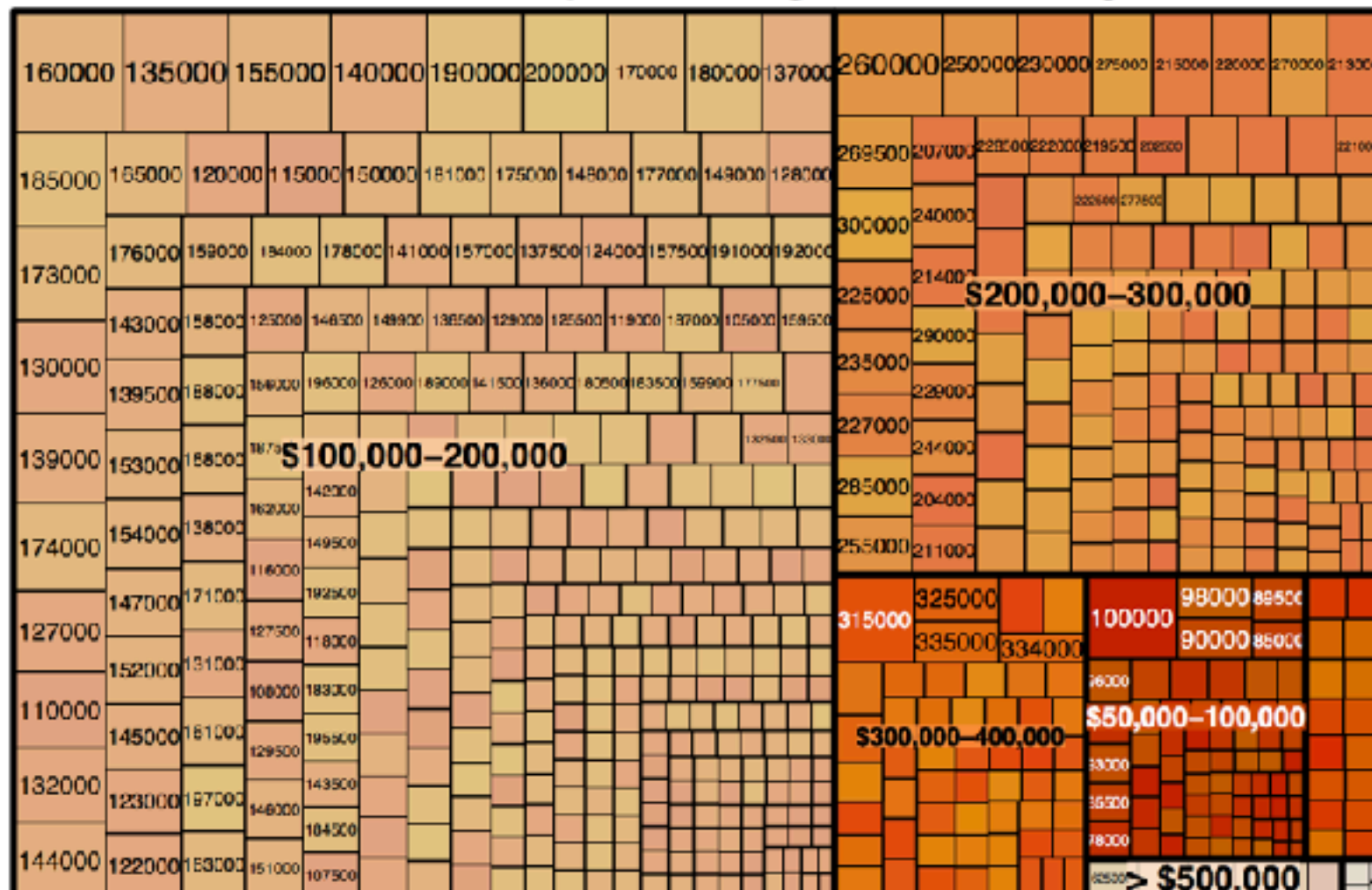
Fig 1: Price by neighborhood

Fig 2: The difference in prices (or lack thereof) between houses close to campus and those further away.

Fig 3: The change in prices with respect to total square footage, stratified by overall quality
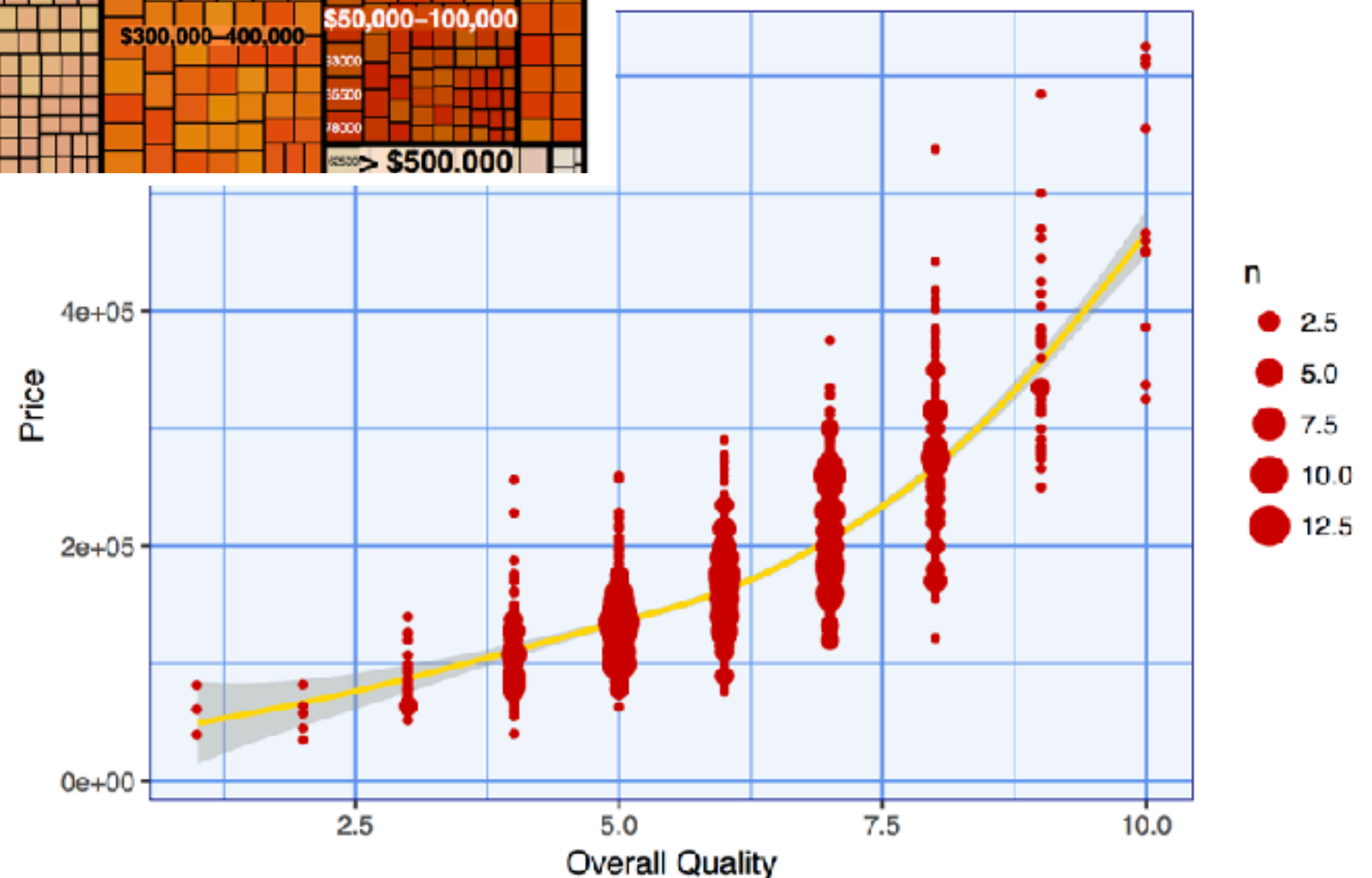
Sale Price versus Total Square Feet by Overall Quality

Plot 5: Tree map of Total Square Footage, Binned by Listed Price.

Plot 6: Price by overall Quality

Methods used to develop simple/complex models

- Scrupulous data cleaning using dplyr and tidyr

- BIC to narrow down the number of variables for simple model

- Boosted tree for relative variable importance

- Pruned decision tree to show us relevant interactions and

- Again BIC to converge on final complex model

bias	max.dev	mean.dev	RMSE	Coverage
-156.1364	64352.67	11229.45	14983.7	0.966

Two Houses to Purchase

PID	price	price_to_pred_ratio	TotalSq	Overall.Qual	Neighborhood
527182020	130000	0.7556536	1204	8	StoneBr
534479120	105000	0.8349797	1376	5	NAmes

- The two houses we recommend purchasing are in the Stone Bridge and North Ames neighborhoods, respectively

- Both in top 10 underpriced houses; both below the median housing price for that neighborhood

- Both of reasonable size in square footage, falling in the middle to low end, relatively speaking

- Overall Quality of 8 and 5, respectively. Could see a big uptick in price with some remodeling to the latter, as we saw in our graph that this uptick generally happens around the 7.5 quality mark.

Insights into Predicting Sales Price

- Despite our initial beliefs: proximity to Iowa State (main employer in town) did not have a significant influence on price