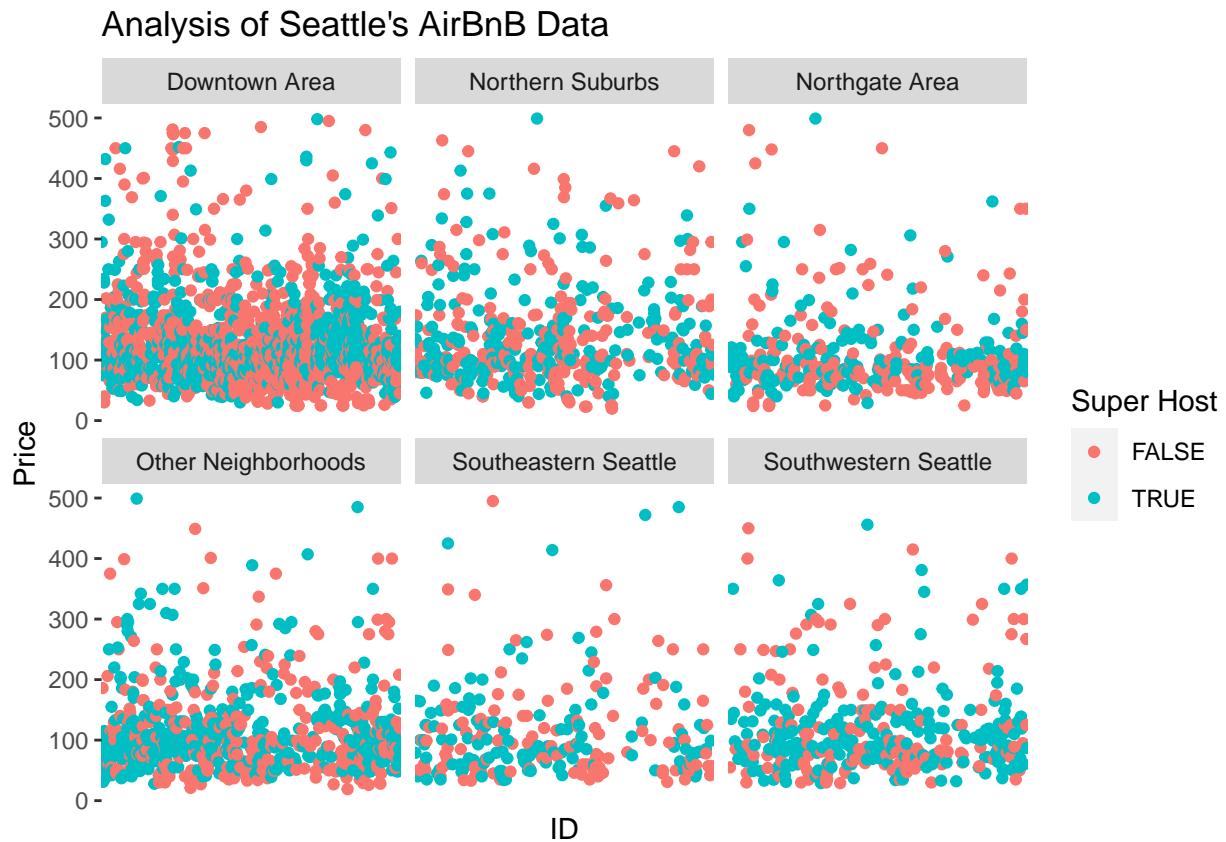


Analysis of Seattle's AirBnB Data

Boston University GRS MA678

Ryan O'Dea

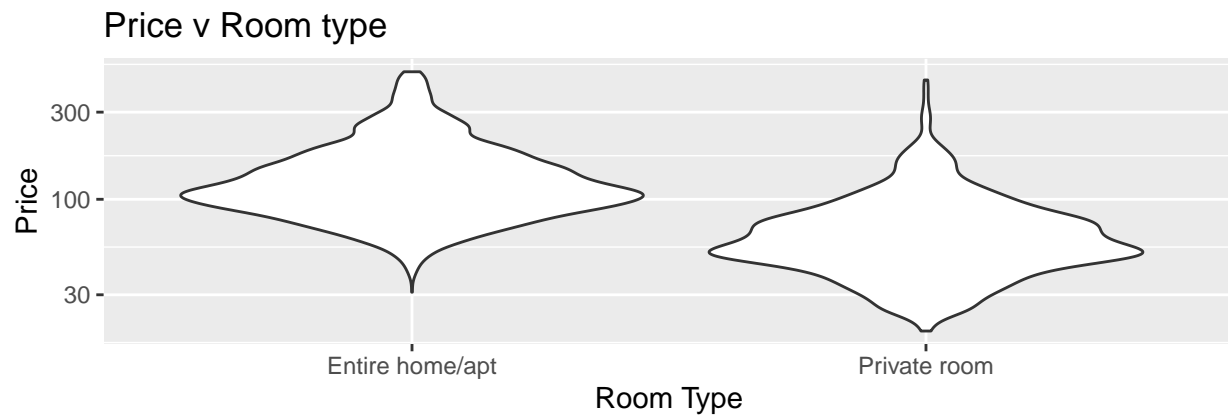
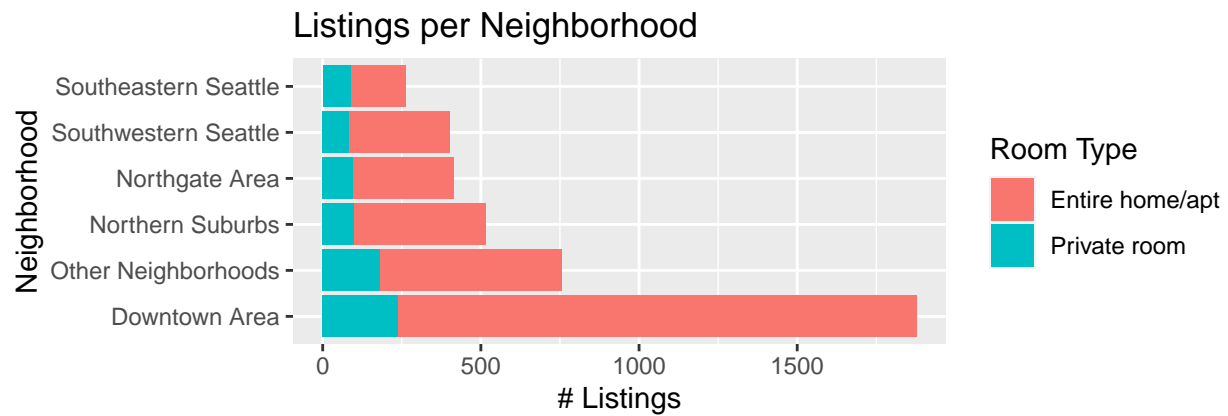
Abstract



Begun in August of 2008, AirBnB is an American vacation rental marketplace. Without owning any of the real estate listings, AirBnB connects hosts, users that are willing to share their home with guests for payment. This project seeks to understand the relationship between how a guest will price their home or single room in the Seattle area compared to factors of location, type of room being offered, and if the host is considered a Super Host, someone who AirBnB has designated as providing a “shining example for other hosts.”

Basic EDA

Introductory Analysis

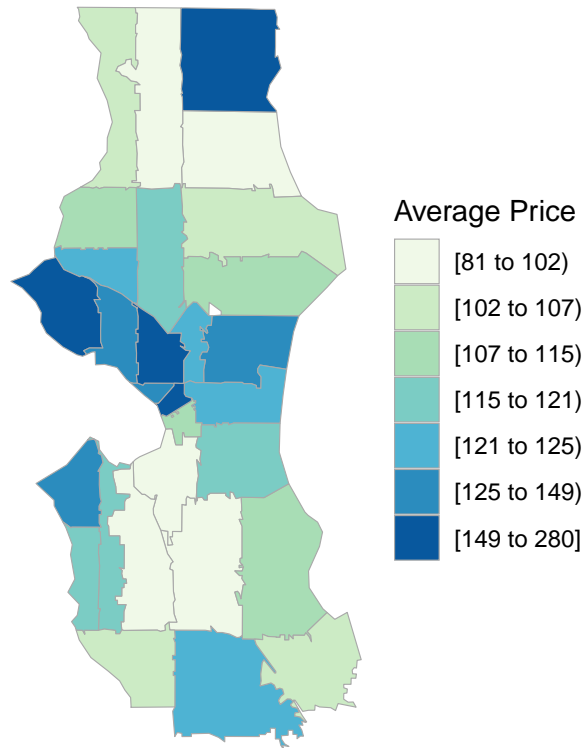


With the beginning exploratory data analysis, we can observe that the majority of rooms are in the Downtown Area and renting an entire home/apartment is generally more expensive than renting a private room with the average entire home/apt with a mean of \$136 per night while private rooms are about half with a mean of \$68.

Spatial Analysis EDA

As expected, the location of the home also plays an important role in determining the price per night. Sorting by zipcode, the downtown area is generally more expensive than the other areas in King Country. We also Lake City (categorized into the "Other Neighborhoods category) in the north falls into the high average price as an outlier for zip codes in the north.

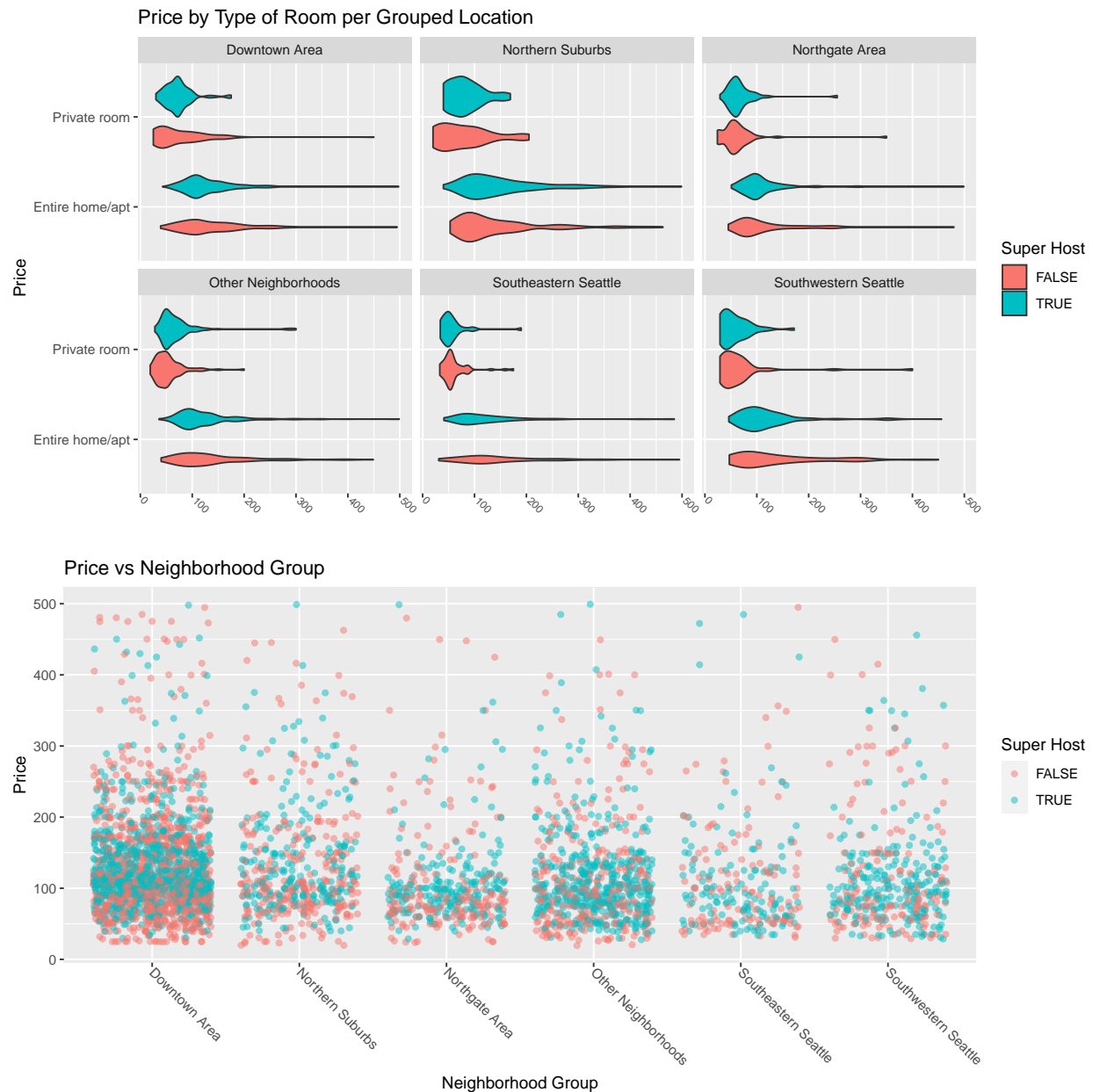
Expensiveness of Seattle's AirBnB's



Fitting a Model

Relation Between Superhost

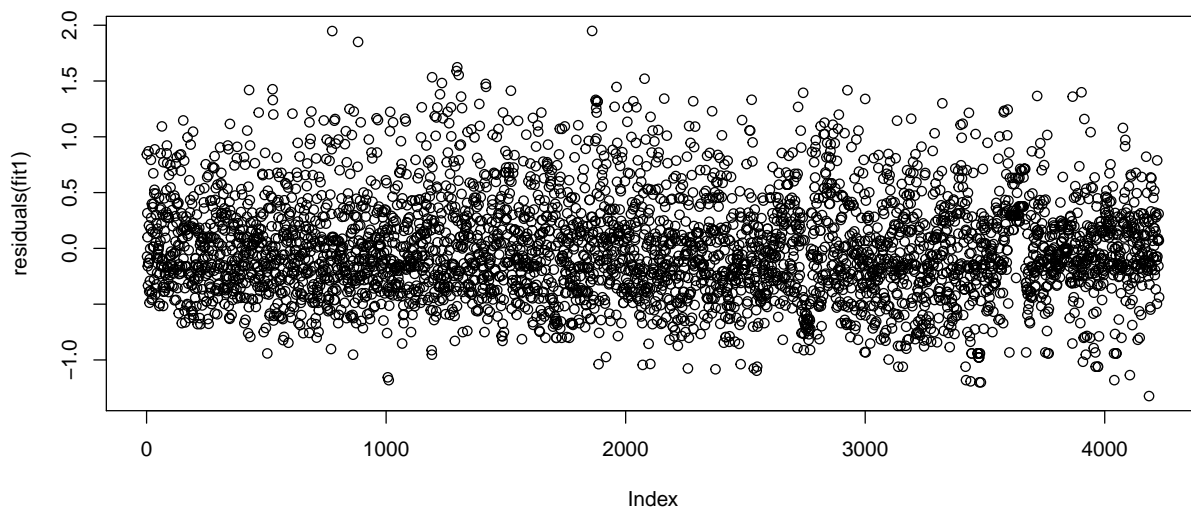
Comparing points and violin plots, there is no apparent relation between being a Super Host and the price per night of the AirBnB. The violin plots appear to have mostly even means, the point plot also shows Super Hosts are interspersed with non Super Hosts when grouped by neighborhood.



Model Fitting

When fitting a model, the log was taken of the price as a variance stabilizing measure. I chose an additive model which observed neighborhood group and room type as predictors. The fit had moderate residuals and an RMSE of 80 indicating a moderately good fit for how noisy the original data was.

The coefficients show both locations and room type are significant when pricing a room; however, my first model included Super Host status as a predictor - which was removed in favour of a better model fit and because it was found to be insignificant. In an analysis of the coefficients, with a baseline intercept of $\exp(4.8)$ - the average price of a full house/apt in the Downtown Area. The other areas are generally less expensive, with the exception of the northern suburbs which are approximately the same, as seen by the summary, additionally private rooms are approximately $\exp(0.68)$ less than their whole house/apt equivalents. Bootstrapping was done to test the coefficients and yielded that the true coefficients are close to the approximate. The full table of coefficients can be found in the appendix.



Appendix

Coefficients

##				
##	Predictor	Estimate	Std. Error	
##	-----	-----	-----	
##	Intercept	4.844	0.011	
##	Nor Suburb	0.036	0.023	
##	Northgate	-0.153	0.025	
##	Other	-0.079	0.019	
##	Southeast	-0.087	0.030	
##	Southwest	-0.684	0.018	