Feedback from Prof:

* Change Q1 to find a model that predicts price and review score
* Finding interactions can be hard. But you decide to make this a prediction problem then considering interactions is reasonable as a way to improve the predictive value of the model
* For building predictive models, choosing factors based on p-values is not optimal, so I recommend you look into model selection methods (I'll be discussing this a bit in class on Wednesday).
* Averaging results for the same host is important. The unit of analysis for your Question 2 should be the individual host
* For Q1, lack of independence because hosts can have multiple listings is a problem for significance testing but not so much for prediction

Action items:

* Change Q1 to find a prediction model for price and review score
* Take average among duplicate super\_host entries from dataset for Q2 √
* Turn assignment into slides
* Visualization

**Questions for Brian:**

* Approach to find right prediction model - say we use Linear regression
  + Do assumption check first √
  + Ways to find right factors to pass into lm()
    - Adjusted R-squared √
  + Once we agreed on the factors to include into model, the predictive equation is found: Y = B0 + B1\*X + ...
  + How to check accuracy of final equation
* Would switching entirely to predictive models affect the quality of our project (evaluation, etc).
* Comment on Q2 of using Logistic regression