**STA 141C Project Proposal**

* **Group Member**

Ying-Chen Chou (ycchou@ucdavis.edu)

Wenqian Li (wqvli@ucdavis.edu)

* **Project Topic**

Predict House Prices with Advanced Regression Techniques

* **Data Resource**

Data is provided by Kaggle

https://www.kaggle.com/c/house-prices-advanced-regression-techniques/data

* **Proposed Work**

This housing dataset includes 79 explanatory variables describing various aspects of residential homes in Ames, Iowa. We aim to predict the final price of each home.

We will first perform exploratory data analysis to check whether the data needs further cleaning and transformation. During exploratory data analysis, we may find more possible features by methods of feature engineering. Then we will adopt model fitting, model selection and advanced regression techniques, such as random forest and gradient boosting, to build reasonable model for predicting house price. At last, we will apply our model on testing data to test the model accuracy.

* **Difficulties of the proposed work**

The most difficult part and time consuming part will be feature engineering when we try to find new meaning features. In addition, it is also challenging to choose the most accurate model for predicting house price.