Group name: **The Outliers**

(1) Names:
Curtis Easton
Andrew Treptow
Alexander Roh
Tod Toter

(2) Project Title: VINOMETRICS

(3) Description of Dataset:

The dataset contains 12 predictors and that can be used to predict wine quality ratings. It was initially 2 csvs, 1 each for red and white wine, but we combined them into 1 csv and created our factor variable, “wine type”, for red vs white. There are 6,497 rows in the dataset.

Initial modeling of the data suggests that most of the variables significantly affect wine quality including the following: Fixed acidity, Volatile acidity, Residual sugar, Chlorides, Free sulfur dioxide, Total sulfur dioxide, Density, pH, Sulphates, Alcohol, wine type.

(4) Background information of the dataset:

The dataset was obtained from the UCI machine learning repository:

https://archive.ics.uci.edu/dataset/186/wine+quality

Its origin was from northern Portugal and it was built to predict quality of wine based on the predictors above. **Paulo Cortez, António Cerdeira, Fernando Almeida, Telmo Matos, and José Reis** from the University of Minho in Portugal built the dataset and published it in their paper: *"Modeling wine preferences by data mining from physicochemical properties."*

(5) Importance of and Interest in the dataset

We have an interest in this dataset as it predicts a unique response variable that we haven’t seen in class. The predictors have interesting relationships relative to the response, and we are excited to see what transformations we may be able to apply to get a better fit. From a business perspective, this study could help inform winemakers as to the qualities of wine that are likely to make it popular. Scientifically it aids in understanding how the chemical properties of wine influence its taste.