**ST309 Draft Project Outline**

**Group members:** Niclas Hallberg and Vlad Ternovyi

**Tentative project title:**

* Wines and Taste – is wine scoring independent from flavour profile, which qualities would net the highest score dividend, and choosing the best quality wine for your budget

**The problems to be tackled:**

1. Expanding and tidying data, i.e. classifying by reds and whites, cutting outliers.
2. Exploratory Data Analysis to summarise some key features of the data set when presenting it in our paper, i.e. relationship between points and price.
3. Text mining and clustering
   1. Possible refinements to flavour profile characterisation, e.g. determining principal flavour characteristics, creating summary metrics based on clustering, collapsing clustering to a 2D plot
4. Regression of points on price and cluster dummies
   1. Possible refinements to regression by using more detailed flavour metrics produced in *3.*
5. Extra (if we have space or the analysis leads us towards these inquiries)
   1. Prediction based on text – of reviewer, country, variety, points, etc.

**The data set to be analysed:** 130k wine reviews with variety, location, winery, price, and description

**The statistical learning procedure(s) to be applied:**

* Text Mining
* Sentiment Analysis
* Clustering
* OLS
* Trees