## Proposal Nick Stracke and Anouk Prins

We will be doing the “Problem” assignment and we will be looking at a dataset from Kaggle.com. It includes data from bike rentals and we want to predict the bike rentals on every 20th of the month. The features consists of information about the weather, what time of year it is, etc. The link: <https://www.kaggle.com/c/bike-sharing-demand> We will be using linear regression and a random forest for this. Our main research question will be: will a linear regression model fit our data better than a random forest?

Literature regarding our research question:

* Variable Importance Assessment in Regression: Linear Regression versus Random Forest, Ulrike Grömping, *The American Statistician*, Vol. 63, No. 4 (NOVEMBER 2009), pp. 308-319
* Classiﬁcation and Regression by randomForest, Andy Liaw and Matthew Wiener, Vol2/3, December 2002