**Dataset: NCAA**  
Source: <https://www.kaggle.com/c/march-machine-learning-mania-2016/data>

**Problem:** Using NCAA from 2003 to 2015, predict what the 2016 Southern conference bracket will look like.

**Hypothesis:** Teams with higher number of historical wins have a higher probability of winning the conference.

**Exploratory:**   
 1. Look at collinearity between the various variables so we can strip it out.   
 2. Use KNN to help predict winners and losers  
 3. Can use grid search to determine which are the strongest predictors  
 4. Use the Sigmoid model to predict likelihood of winners and losers  
 5. Use True Positive Rate to test the accuracy of the model  
 6. Use cross validation to test my model.

**Dataset:** the data is available through Kaggle, although I would have to link the various data tables together

**Features:**  
 1. % of field goals missed  
 2. % of three pointers made  
 3. Number of offensive rebounds  
 4. Number of defensive rebounds  
 5. Day of the game  
 6. Strong or weak seed

**Goals/Criteria for success:** To be able to predict a bracket that at least will predict the correct winner for the southern conference.