Phase 3 - Team NFL - John Monahan, Pete McCoy, Eric Libby, Nicholas Pickett

Analysis Plan

The data that we are submitting for phase 3 is not yet complete, as we are still waiting for one data set to be completed (we each took one year).  After we get all four, we may do more hunting for combine statistics elsewhere in order to get a complete dataset. We need to see what combine drills have the most accessible data overall.  Also, we may do some type of system in which we assign numerical numbers to each big college, so we can do a representation of what colleges produce the most NFL players in the first three rounds. A good visualization technique to understand where successful college players come from could be clustering. Per our class discussion, we can write some code to create a graph with a certain amount of clusters. This will give us the ability to make a prediction on where the successful players come from and help us with our final prediction of overall success from college ball to the NFL.

Another change we will make is the column named “To” in the data(which states what year they played until), will be changed to the years played as a numerical number 1-10+.  We are doing 10+ because some players are still active, and anything over 10 years is a substantially successful career.  We are waiting to make some changes until we have all of our data together and ready to go.

We will partition the data by using two years for the training sets and another two years for the test set.  We have not yet decided what years will be for each, and we may even do three training sets and just one test set.  One visual that we hope to include that was previously mentioned is a representation of what colleges produce the most players in the NFL.  We can also see which colleges produce the most successful players.  These two are things that we just thought would be interesting to see and could make an interesting visual.  The main analysis that we will perform is based off of the question of- does college statistics and combine results give any insight on an NFL career?  We also may use classification and different types of regression.  To do this, we will plug in different variables that we think are important to NFL success, and see if the model can find any patterns within.  Also, using JMP like we did in last class could be another possible way of going about our analyses.  It gives much better visuals and has some different things that we could work within the clustering commands. We will be able to do this analysis with career length, pro bowls, and all-pro awards; so we have some flexibility with what target variable we will use. I am sure we will come up with more analyses, but we have yet to gather together to collaborate and brainstorm the ideas.