Project proposal
Advanced Methods III 140.753

As a project for this class, 140.753, I'm interested in learning more about the statistics methods and concepts that have been applied to both NFL and NBA players[1]. I believe that others have tried to de-mistify the games and explain numerically what others justify with intuition. One great example is what Brian Burke has been doing at Advanced NFL Stats, including a compilation of play-by-play data for the last 10 years at `http://www.advancednflstats.com/2010/04/play-by-play-data.html`. On a more sparce level, Rafael Irizarry has made some interesting points regarding basketball players at `http://simplystatistics.org/tag/basketball/` including his latest piece regarding how Kobe's ball hogging affects the Lakers.

As we saw in class [lecture 003], a good project leads with the question. At first I thought of trying to explain what made the Baltimore Ravens win over the New England Patriots at the AFC championship of the 2012 season so special in terms of reflecting the so-called *intangibles*. My other main interest is comparing the characteristics of the game at half time versus all other half-time games from the Patriots and other teams. After all, was it really special that the Patriots were 67-0 when leading at half time? If I had to guess, the facts that it was a low scoring game half way through with a small difference in points surely gave the Ravens a strong chance of winning.

I have to admit that my knowledge of sports sabermetrics is quite limited. For instance, after reading What Makes Teams Win? [4 part story], I think that the effect of the *intangibles* is going to be very hard to describe. After all, Brian Burke did an analysis of Best Games of 2012 and Best Playoff Games using what he has defined as the *excitement index* (EI) and the *comeback factor* (CBF) which showed that the AFC championship was not so special as can been seen in `http://wp.advancednflstats.com/nflarchive.php?gameid=55834`. To further highlight Brian Burke, I liked his new piece on Who's the Clutchiest Post-Season QB? where I was fairly surprised by Brett Favre's (obviously one of the best QB ever!!) outlier position (see the *All Years* tab). This piece also highlights to me why it is important to make interactive graphs for this kind of results. Anyone looking at it, and with some knowledge of the game, will want to find their player of interest or change some parameters like the year in question.

I'm aware that previously other students have played around with sports data. Namely, Lei Huang, Prasad Patil and Therri Usher. Therefore, my project proposal is the following one.

- Further understand current sabermetrics knowledge. For example, the *Winning Probability* and *Expected Points Added* which are so heavily used by Brian Burke.
- Learn what other students tried in the past and learn from their experiences.
- Define a specific question of interest.
- Get going with the project! Aka, gather data, pre-process, explore, analyze, conclude.
- Make an interactive report using (if possible) some of the latest tools such as the Shiny web application server made by the people at RStudio. Preferrably, using ggplot2[2] as the graphics device (it is possible) to make the plots look neat.

[1] I figured that the project should have nothing to do wiht our research to promote that *you can do it!* attitude.
[2] I'm interested in learning how to use ggplot2 and have ordered 2 books on the subject. Seems like it already displaced `lattice`.