Vanessa Ohta

October 13, 2015

Data Science Final Project Proposal

I want to do a project predicting NFL player performance (most likely limited to QBs) based on their Twitter activity. I started thinking about this after reading a couple articles that just look at descriptive statistics of athlete’s tweets and their performance ([Russell Wilson should tweet less before games](http://www.si.com/extra-mustard/2015/09/14/russell-wilson-twitter-stats-seattle-seahawks) and [The Great J.R. Smith Sleep Study](https://storify.com/MylesMaNJ/the-great-j-r-smith-sleep-study)). I think that it will be interesting to look at the sentiment of their tweets in addition to the frequency/timing of their tweets. Looking more into this, I found several other people that are looking at Twitter sentiment’s correlation with team performance as well.

There are two sets of data I will need for this:

1. NFL Performance Data: I found [this API](https://github.com/BurntSushi/nflgame) to retrieve and read NFL Game Center data, but I haven’t explored it in depth. It looks like it will have all necessary player statistics including those that are used to calculate passer ratings.
2. Player twitter data: I have not yet begun collecting this data, but can pull from the Twitter API. I do not know very much about it yet!

I’m interested in this topic, because I enjoy Twitter and the NFL and I think that it will be fun. I will also have to use several different techniques that will be new to me in getting, cleaning, and analyzing the data, which I think will make this a good learning experience for me.