Here is my feedback on your project proposal. Your grade should appear in AutoLab soon, but it depends on when the other proposals get graded. We will normalize by the grades that the other TAs give, so this isn't your final score.

Your project sounds very cool, and I think it's an interesting question. I also think you have very carefully thought out your plan of attack, which is good. For future reports, you need to proofread your writing a little bit more - there are many grammatical mistakes.

I am concerned about two things:

1. You may get behind because you have so much data collection/preprocessing to do. If you are on track with the schedule you propose then it's fine, but data preprocessing can sometimes take much longer than expected. The focus of the project should be machine learning, so don't spend too much time on this. Let me know if it is proving to be too difficult and we can try to discuss workarounds.

2. The question you want to answer (based on my interpretation) is what makes a good team, correct? Your formulation is to have team composition represented as a feature vector, labelled with the result of the match that team played in. I don't see how the proposed methods (SVM/LogReg) can answer your question better than simply correlating the feature vector with the match results. It seems like what you want to do is not just build a model that predicts game outcomes, but also *interpret* that model so that you can make claims about the feature space.

If my assessment is correct, then you may want to look at machine learning techniques designed to answer questions about which features are highly predictive. One way to do that is with feature selection via lasso or group lasso. A challenge when using lasso is that if there are sets of features that are highly correlated (for example, how many games a player has played is correlated with how many games he/she has played as a particular character), it will tend to give noisy results. You can look into overcoming this challenge.

This project should be about more than just throwing pre-packaged machine learning algorithms at the data set. Given that you don't have a state-of-the-art to compare to (although you may want to see literature about sports team analysis), how can you use domain knowledge of the feature space to pick/make smarter algorithms to answer your question?

Question: 5/5
Data set/Evaluation: 3/5
First Steps: 4/5
Goals: 4/4
Neatness: 0.5/1
Total: 16.5/20

If you want to meet and discuss your project, then please come to my office hour, and let me know beforehand.
All the best,
Nicole