Group Paper

Data Programming in R

Group O - Sam Anderson, David Sorensen, Will Cobian-Ruelas

Accuracy of Rankers for College Football Recruiting

**Introduction**

**Getting the Data**

**Cleaning the Data**

**Functions**

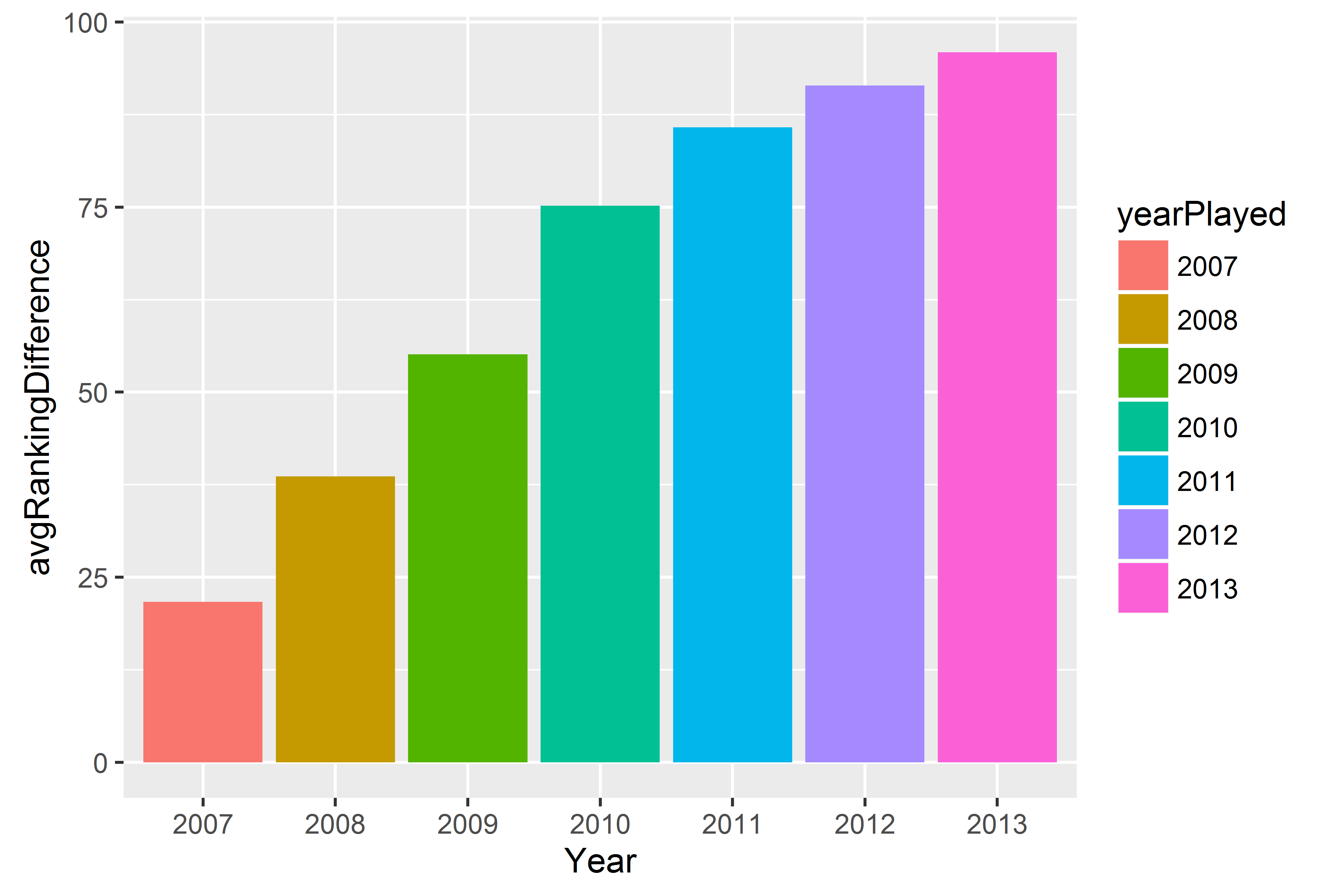
**Problems**

**Outcome**

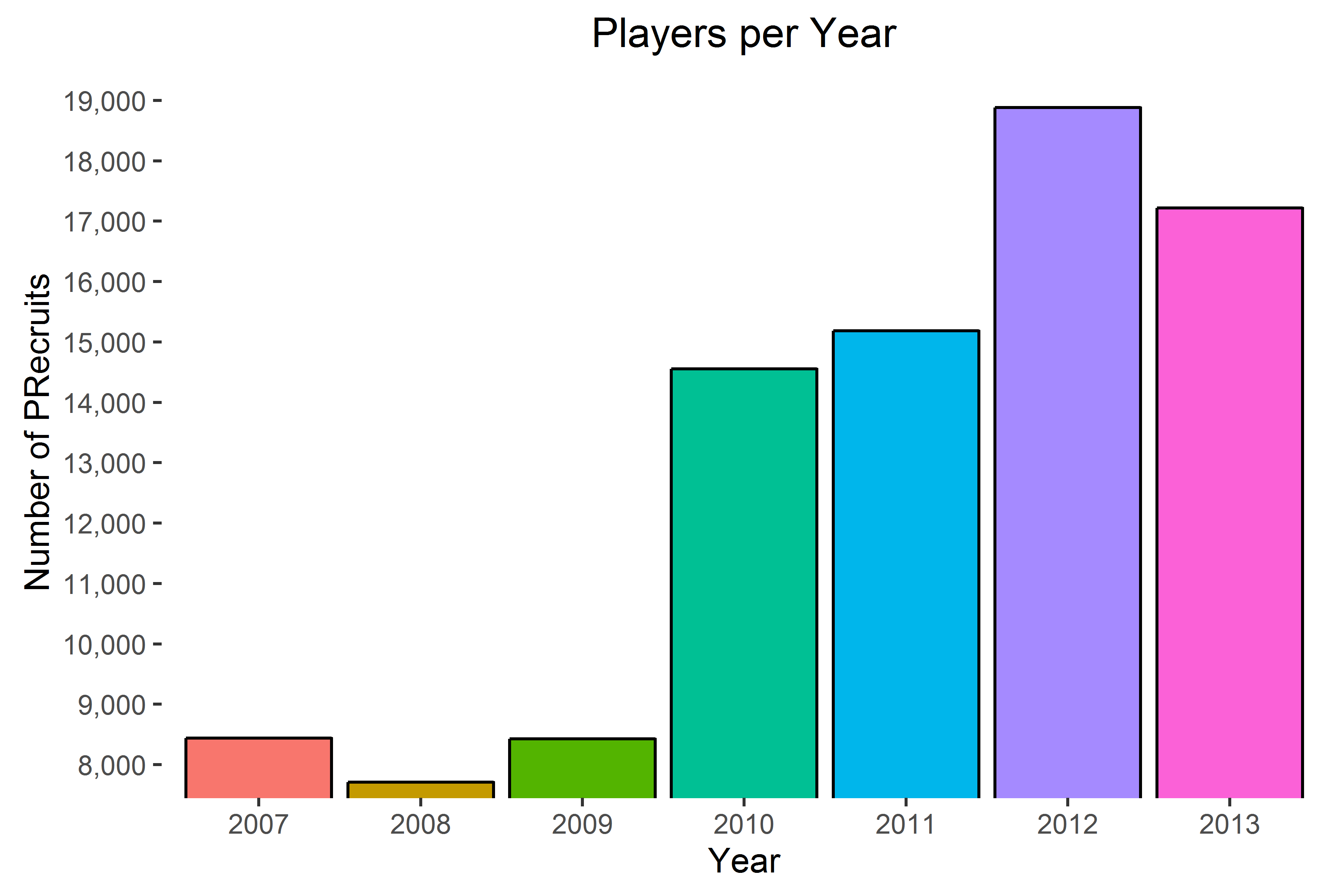
To determine the accuracy of rankers for college recruiting, we took the difference between a players ranking out of high school and compared it to their ranking after each college season that they played. We then took the average difference by position and year to determine how accurate the rankers turned out to be. Note here that we used difference instead of variance. To determine the average, we need to be working with a positive number, but variance was giving us how far off the rankers were in a positive and negative direction (i.e. whether they over-ranked or under-ranked each of the prospects). So we defined the difference to be the absolute value of this difference so the positive and negative numbers would not cancel each other out when we calculated the average.

The results of our study can be broken down into two different situations based on how we decided to deal with NA value in the data pulled from ESPN. Our two different handlings of the NA data results in two different interpretations of the data. I will go into both ways we handled NA below as well as what we concluded from the results.

The first time through, we replaced all of the NA grades with a 49. We settled on a grade of 49 because that was one point lower than the lowest graded prospect. In our minds, if the ranker decided not to grade a recruit, then there was no difference between them and anyone else with an NA grade. The reasoning behind using a 49 to replace NA was sound at the beginning of our analysis, but we saw flaws with it when we started to analyze our results. From 2007 to 2013, we saw the average difference in rankings increase greatly (see chart below).



Initially, we concluded that rankers were getting worse as the years passed. But then we looked at the data further and found what might actually be the cause of this increase. From 2007 to 2013, the number of recruits per year in our data also increase greatly (see chart below). Recruiters were still ranking approximately the same number of recruits, so the number of NA’s per year increase as a result. Some of these NA’s produced at a high level when they played in college, so it increased the average difference in rankings per a year. This result is really a reflection of the increase in the number of high school recruits trying to get ranked, and it is not a good measure of how accurate recruiters are in their rankings. It just shows us that there are too many players out there to be ranked every year. Because of this determination, we decided to handle the NA values differently and come to a real conclusion about the rankers.



Our second time around, we decided to drop any recruits with an NA rank from our data. This resulted in a much better measure of the accuracy of the rankers. We didn’t the see the steep decline in year-to-year accuracy that we saw in our first attempt. We also saw accuracy overall get much better than what we saw when we replaced the NA grades with a 49. Our new positional and yearly charts (see below) now had a difference of about 30 spots as a ceiling, down from the nearly 100 spot difference we were seeing before. This allowed us to conclude that when a player is ranked by the experts, they are actually pretty good about determining future output at the collegiate level. They were better at some positions than others, but overall they did a pretty good job. We would be comfortable using the expert’s rankings as a solid tool for determining whether or not a player will have success as a college football player.

