210 Project

Austin Brown and Ryan Yu

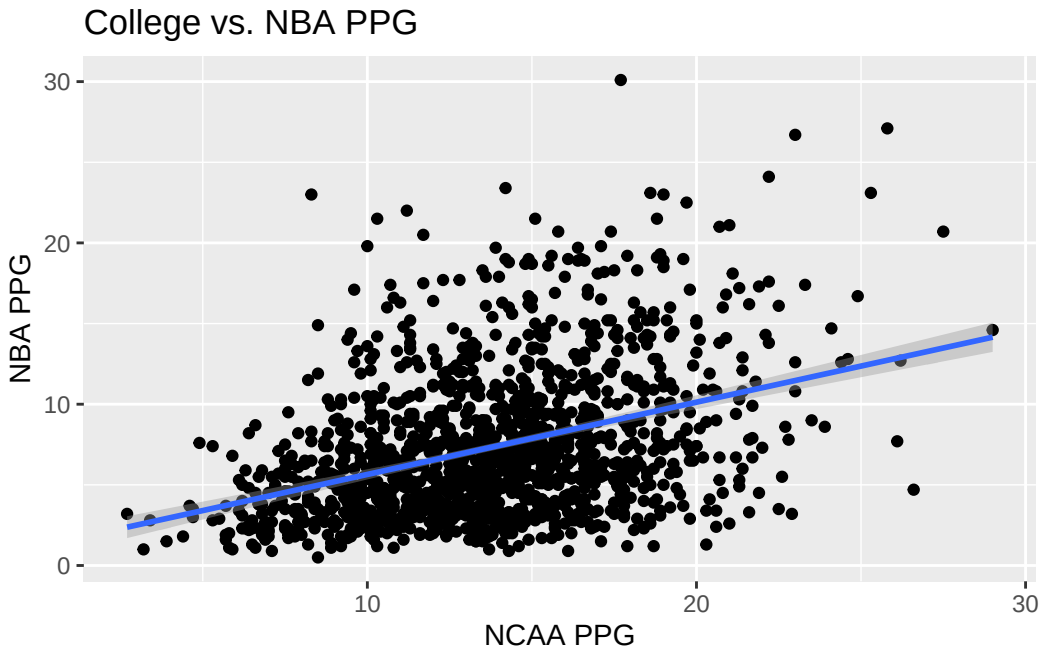
Position and NCAA Career length Variables

```
[1]  16 152
```

Introduction

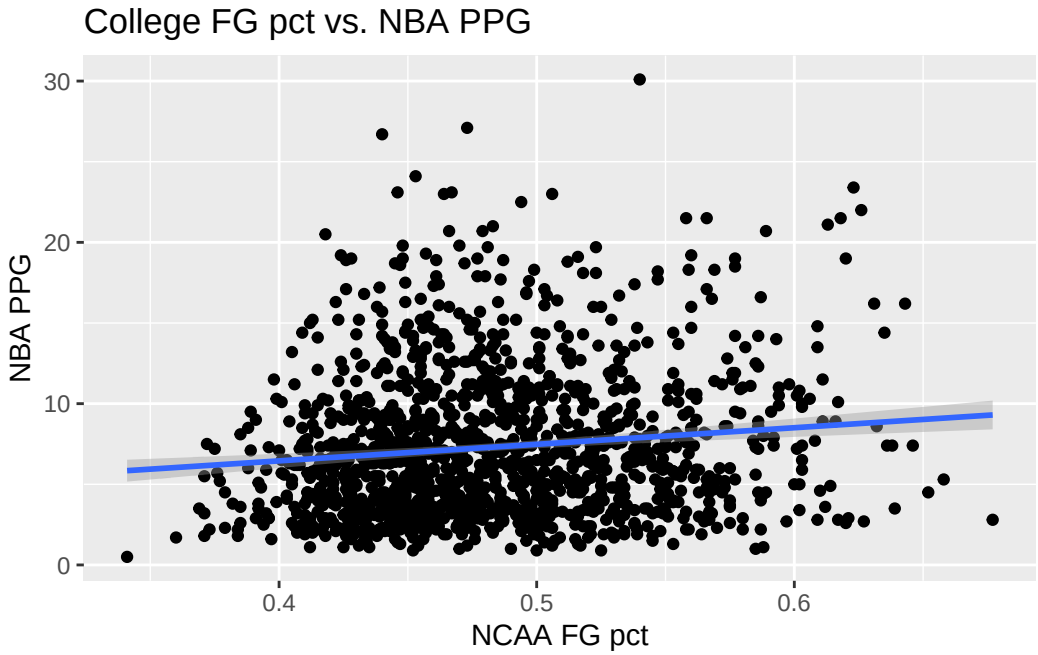
Data

EDA

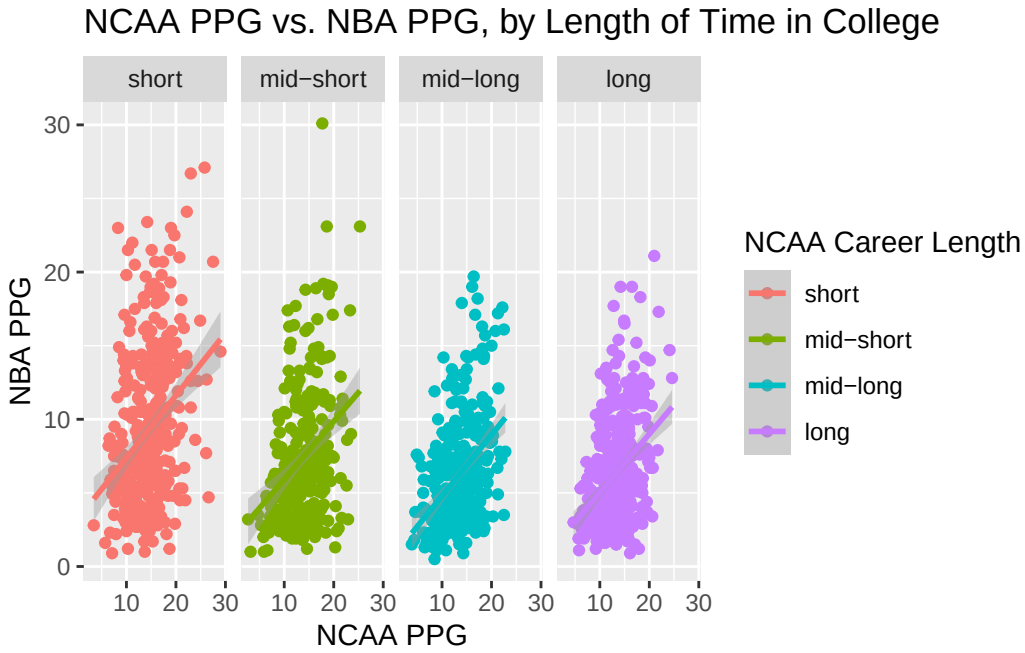


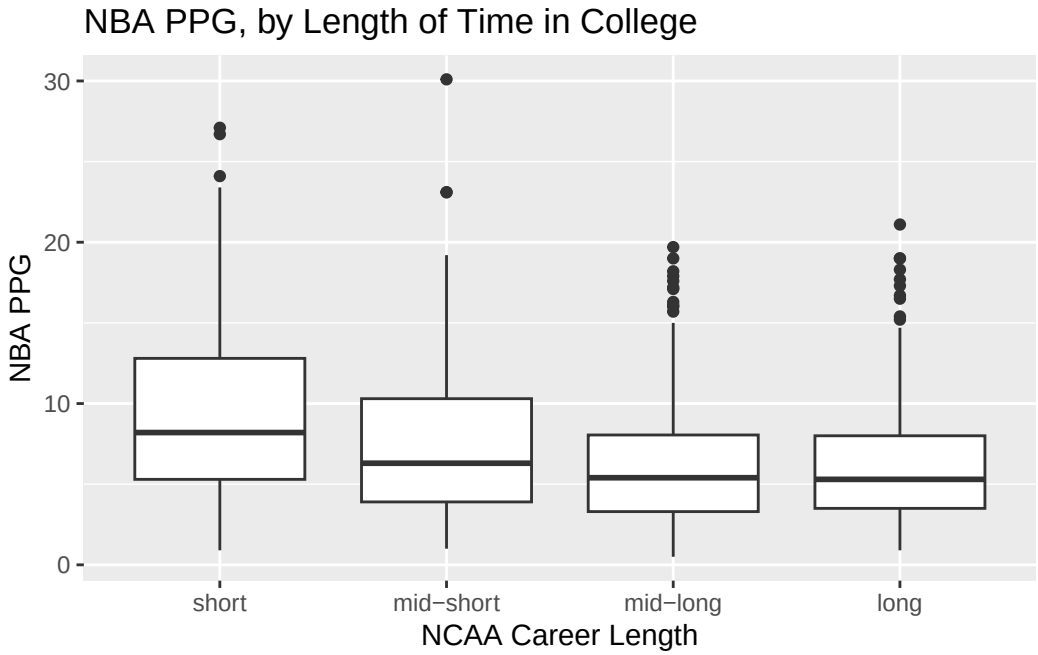
Our first initiative was to compare the direct analog to NBA ppg, which is NCAA ppg. As can be seen in the scatter plot above, there is a direct positive relationship between NCAA

PPG and NBA PPG. One interesting thing to note, however, is that there are cases of 'high' NBA scoring averages (> 15) even for players who had NCAA scoring averages of less than 15 PPG.



The second relationship we wanted to explore is between scoring efficiency in college (in this case NCAA FG pct) and NBA PPG. We also see a generally positive relationship between NCAA FG pct and NBA PPG, however, the fitted line in this plot has a distinctly less extreme slope than in the previous plot. Both of these plots suggest a relationship between scoring ability and efficiency in college and NBA, however, it is likely that there are more factors involved in predicting. especially given that at the college level, most players may not be fully developed from a skill, physical, and mental standpoint. We explore this further in the next two graphs.

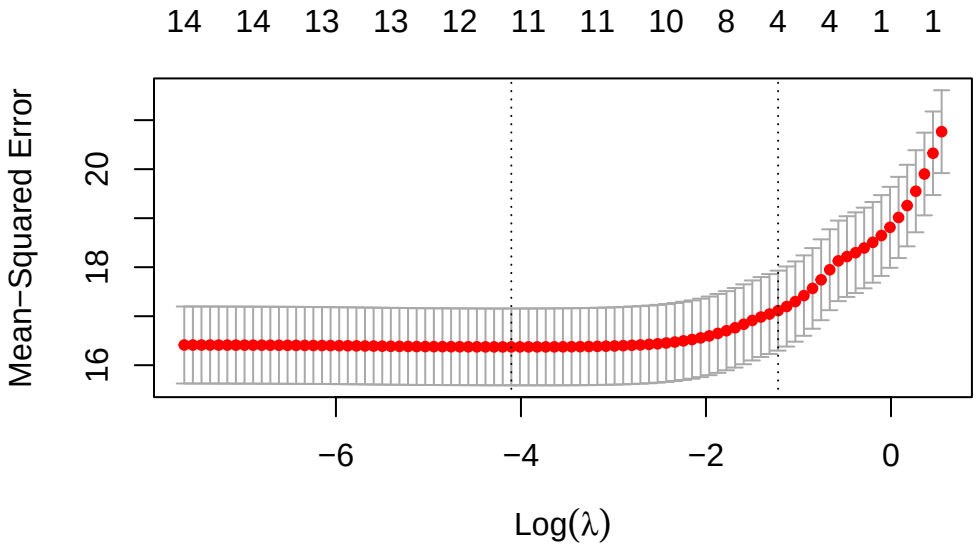




In the two graphs above, we wanted to get a sense of NBA scoring on a college career length basis. What we can see from both plots is that the two shortest categories for college career (likely players that left early for the NBA) have observations that spread higher in the NBA ppg category. All four categories maintain the upward sloping fitted line in the scatterplot,

however, there is a clear difference amongst the four categories in the NBA scoring averge distributions. This is likely due to the fact that players who have 'lottery pick' potential for the NBA draft will often leave college early. These players are more often the players who develop into the most dangerous scoreres in the NBA. As we continue to explore these relationships and begin to craft a model, college career length will certainly be taken into consideration.

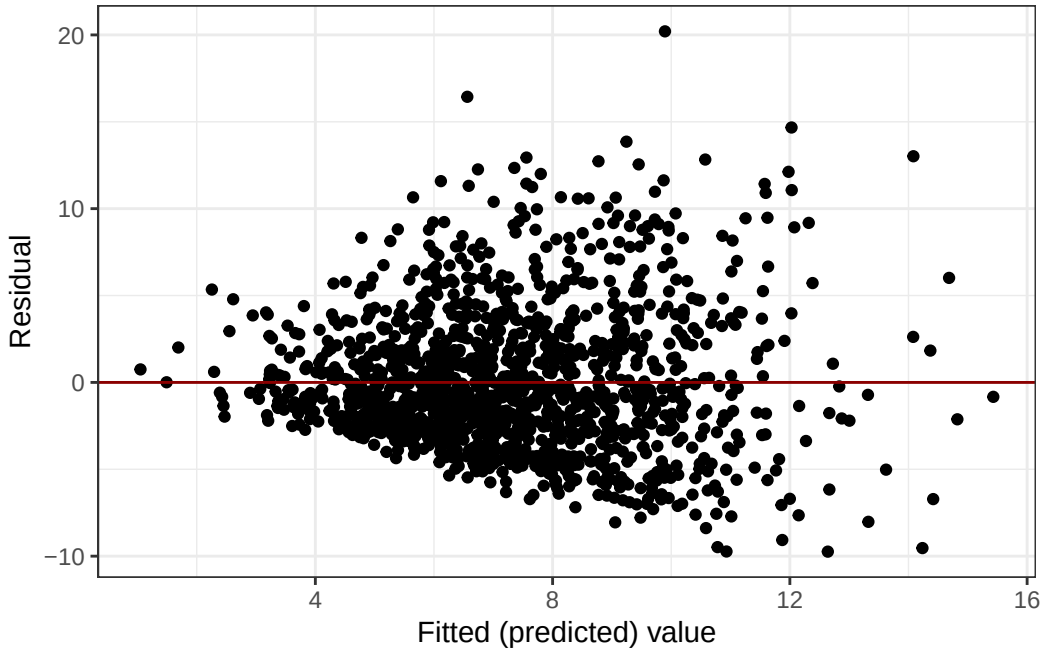
Methodology

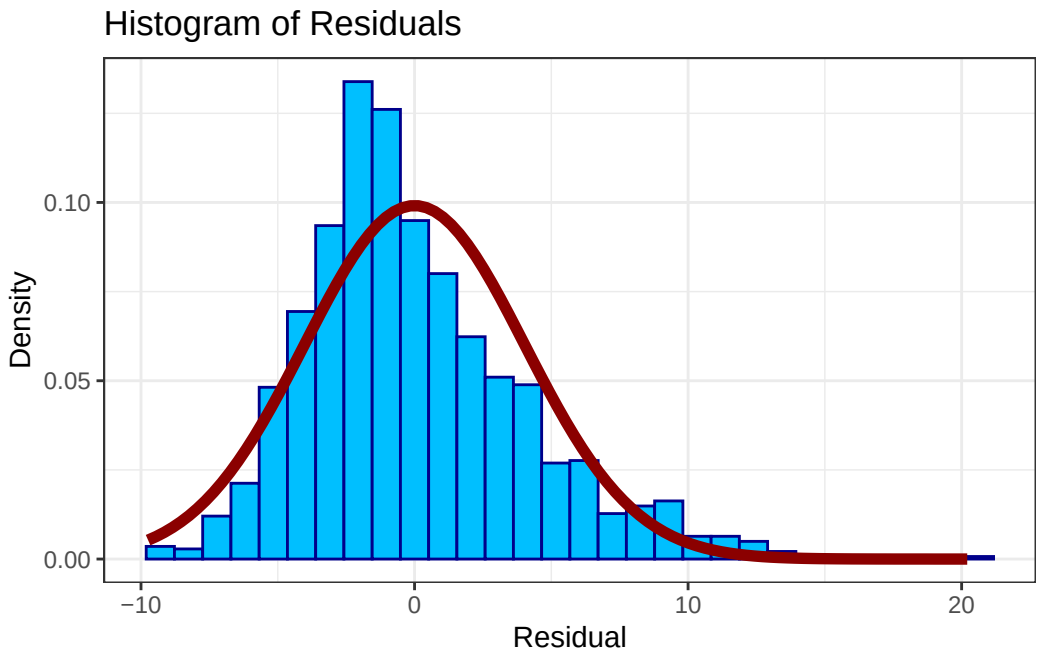


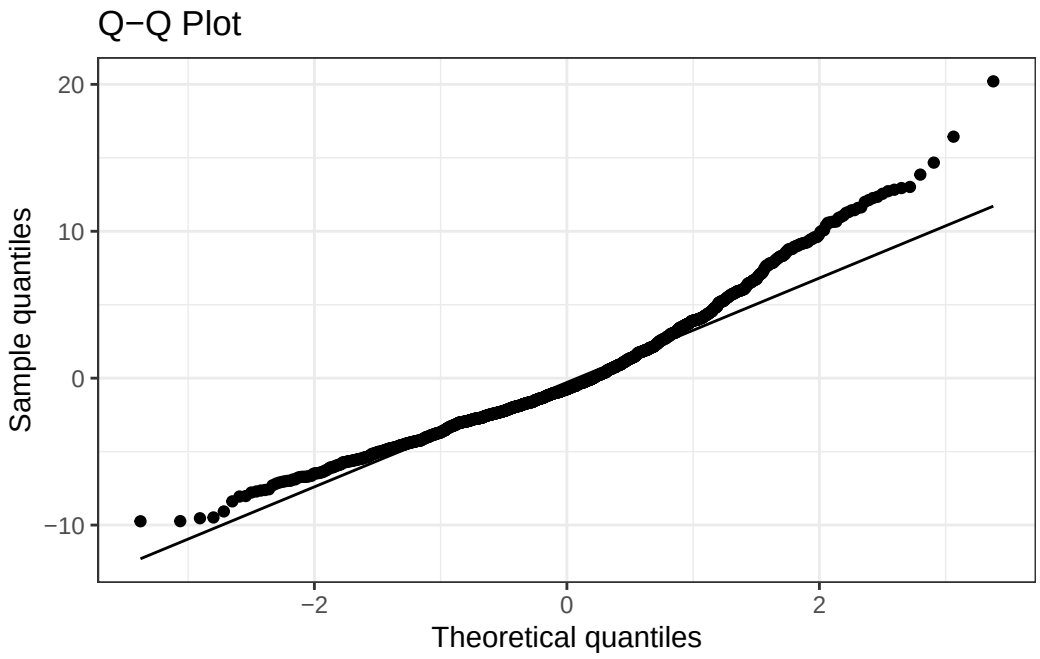

```
15 x 1 sparse Matrix of class "dgCMatrix"
                                       s0
(Intercept)                  .
NCAA_ppg                     0.389509975
NCAA_ft                      1.665822912
NCAA__3ptpct                 1.512781668
NCAA_fgpct                  15.292720454
NCAA_lengthmid-short        -1.226583839
NCAA_lengthmid-long         -2.481722769
NCAA_lengthlong             -2.032078068
positionF                   -0.038877682
positionG                    0.670013209
NCAA_ft:positionF            .
```

```
NCAA_ft:positionG                .
NCAA_ppg:NCAA_lengthmid-short -0.021265024
NCAA_ppg:NCAA_lengthmid-long  -0.006795547
NCAA_ppg:NCAA_lengthlong      -0.037416291
```

Assumptions Evaluation







Results

```
# A tibble: 11 x 5
```

	term	estimate	std.error	statistic	p.value
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
1	(Intercept)	-5.06	1.86	-2.72	6.58e- 3
2	NCAA_ppg	0.398	0.0534	7.46	1.59e-13
3	NCAA_ft	2.63	1.68	1.56	1.18e- 1
4	NCAA__3ptpct	1.95	0.935	2.09	3.72e- 2
5	NCAA_fgpct	12.6	2.19	5.74	1.15e- 8
6	NCAA_lengthmid-short	-1.15	1.18	-0.980	3.27e- 1
7	NCAA_lengthmid-long	-2.57	1.14	-2.26	2.42e- 2
8	NCAA_lengthlong	-2.00	1.12	-1.80	7.25e- 2
9	NCAA_ppg:NCAA_lengthmid-short	-0.0314	0.0798	-0.394	6.94e- 1
10	NCAA_ppg:NCAA_lengthmid-long	-0.00686	0.0787	-0.0871	9.31e- 1
11	NCAA_ppg:NCAA_lengthlong	-0.0462	0.0784	-0.589	5.56e- 1

From this model, based off of p-values and a significance level of .05, we can identify the following terms as significant: NCAA_ppg, NCAA_3ptpct, NCAA_fgpct, and NCAA_lengthmid-long, as all of these terms have p values of less than 0.05. As for interpretations, if NCAA ppg increases by 1 point, we can expect that on average, NBA ppg will increase by 0.398 points while holding all other predictors constant. For NCAA_3ptpct, if that increases by 1, or 100%, we would expect NBA ppg to increase on average by 1.951 points, while holding all other predictors constant. To interpret the percentage terms on a different scale, for NCAA_fgpct, for every 1% increase, we would expect on avergae NBA ppg to increase by 0.126 points while holding all other predictors constant. Lastly, if a player has a 'mid-long' college career, we would on average expect their NBA ppg to be 2.566 lower in comparison to a player with a 'short' college career, while holding all other predictors constant.

In terms of answering our research questions, the interpretation of the first three terms was logical and in line with our expectations. In general, players who score more in college, as well as are more efficient both from the field overall but also from three point range are likely to be better scorers in the NBA. The last term we interpret is interesting, as it is suggesting that players who stay longer in college are expected on average to score less than the shortest college career length. This can be explained by the 'one and done' phenomenon we briefly mentioned earlier, in that the players with the most potential NBA upside usually leave college early, even if they haven't fully developed yet as a player. In comparison, players who stay in college longer may have less natural scoring ability, and thus on average may score less in their NBA careers.

To check the predictive power of our model, we are going to compare our model to a baseline model, that uses college points per game as the sole predictor as one of our goals was to evaluate whether NBA General Managers should evaluate player scoring ability with more than just their college points per game.

Baseline Model Results:

```
     RMSE  Rsquared      MAE
1 4.21319 0.1506385 3.256958
```

Experimental Model Results:

```
      RMSE  Rsquared      MAE
1 4.046686 0.2173287 3.136414
```

Discussion and Conclusion

Appendix