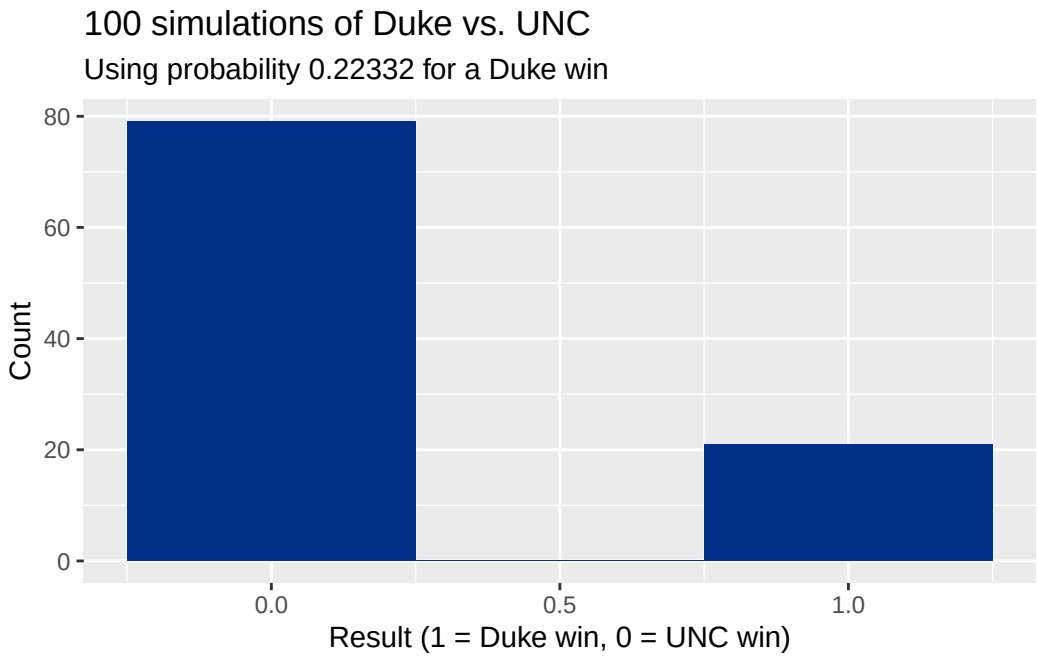
Predicting Duke vs. UNC Feb. 3, 2023 using modeling

Dom Fenoglio, The Duke Chronicle

What is the game prediction?

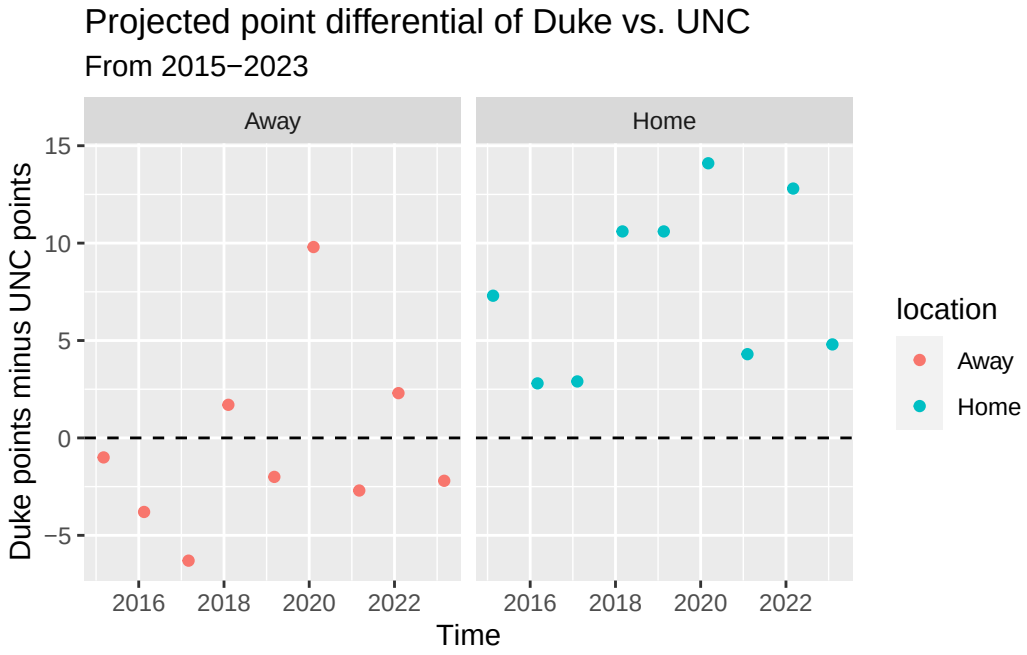
```
        date team            opp game_location    tempo      ppp  pts  win_per
1 2024-02-03 Duke North Carolina             A 71.22449 1.017415 72.5 22.33198
  did_win simulate_date year
1   FALSE    2024-02-03 2024
```

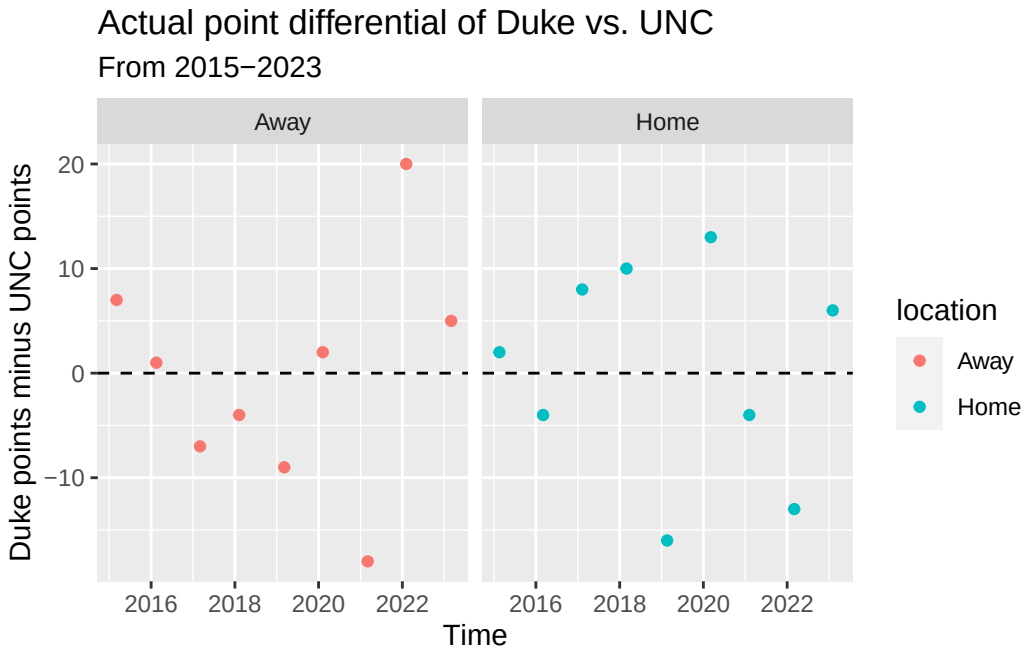
```
        date           team  opp game_location    tempo      ppp  pts  win_per
1 2024-02-03 North Carolina Duke             H 71.22449 1.133885 80.8 77.66802
  did_win simulate_date year
1    TRUE    2024-02-03 2024
```



This is a simple representation of 100 simulations of a binomial trial with the given probability of Duke winning. In other words, if the game was played 100 times, these would be the results. Clearly, the Tar Heels are going to run away with this one. But, this is just a surface level look at the data, so let's look closer.

Some exploratory analysis

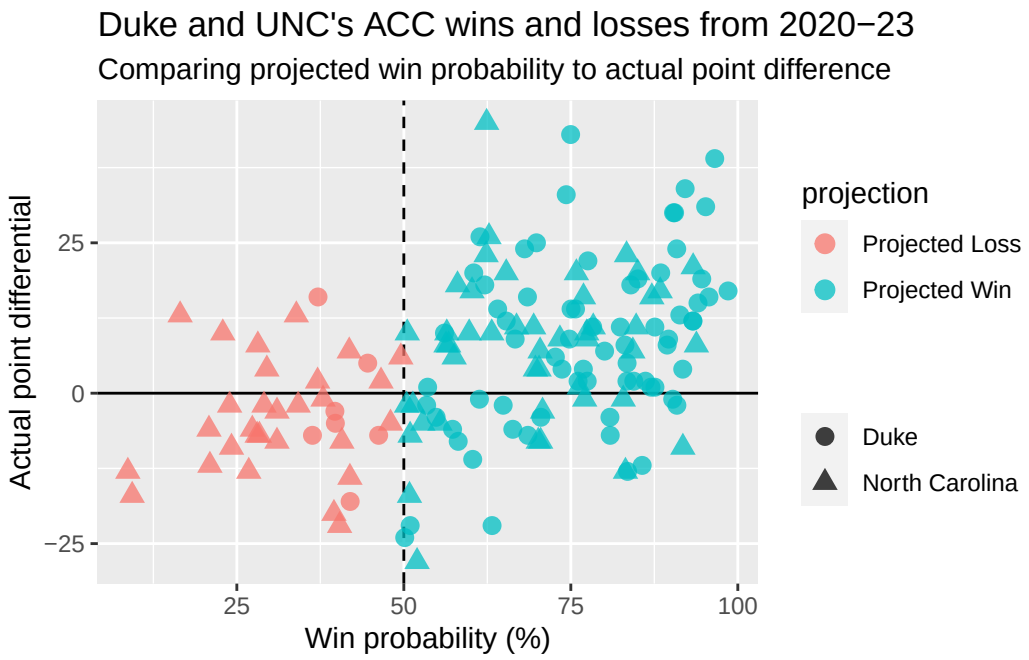




These graphs allow you to see the vast difference between projected point differentials and actual game scores. In particular, notice the blue dot dated right after 2022. This was Coach

K's final home game. An unranked North Carolina team beat No. 9 Duke in a decisive 13-point victory. On the whole, the predictions can get really far off. Maybe it's just hard to predict basketball games (it is), and the model can't do much.

Duke and UNC ACC games 2020-2023

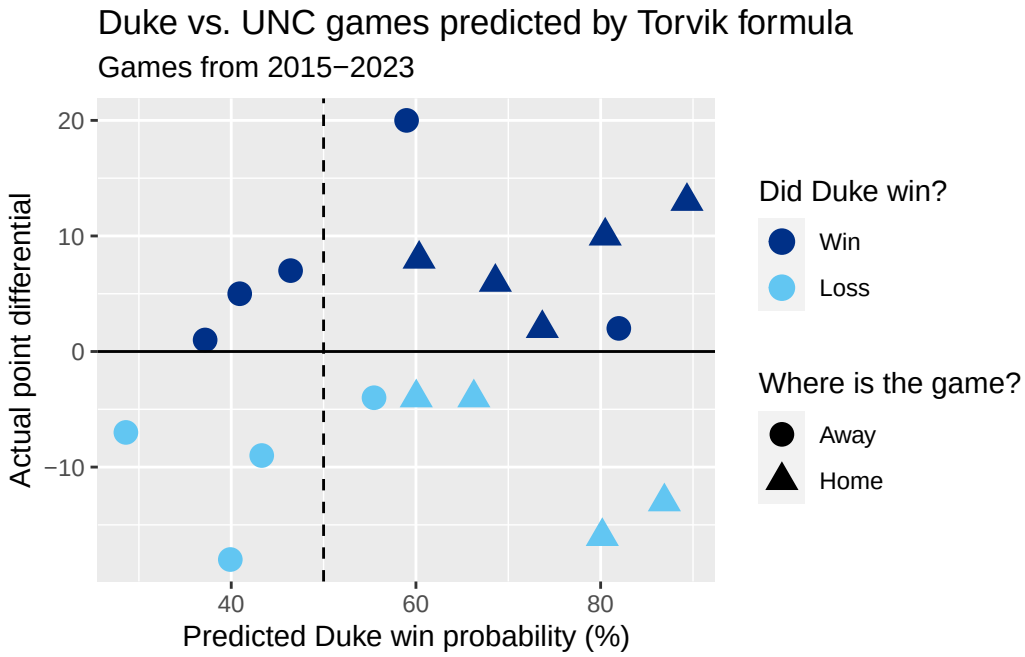



```
Confusion Matrix and Statistics

          Reference
Prediction  0  1
         0 25 11
         1 33 85

               Accuracy : 0.7143
                 95% CI : (0.636, 0.7841)
    No Information Rate : 0.6234
    P-Value [Acc > NIR] : 0.011316

                  Kappa : 0.3421

 Mcnemar's Test P-Value : 0.001546
```

```
            Sensitivity : 0.8854
            Specificity : 0.4310
         Pos Pred Value : 0.7203
         Neg Pred Value : 0.6944
             Prevalence : 0.6234
         Detection Rate : 0.5519
   Detection Prevalence : 0.7662
      Balanced Accuracy : 0.6582

       'Positive' Class : 1
```

This graph allows us to see the distribution of Duke and North Carolina's ACC games from the past few years, and how the model did at predicting the outcome. On the whole, it did a pretty good job, as shown by the confusion matrix analysis. Our confusion matrix gave a significant p-value of 0.001546, a positive predictive of 72.03% and a negative predictive value of 69.44%. Great! Then, it should be able to be just effective at predicting Duke vs. UNC, right?

Now, let's see the past 8 years of Duke vs. UNC



This graph breaks down win probability versus actual results, with coloring showing which shade of blue won. A quick look at this graph shows that our model didn't do a very great job. Why are these games so far off if on the whole it does a good job? Again, let's look closer.

```
Confusion Matrix and Statistics

          Reference
Prediction 0 1
         0 3 3
         1 5 7

               Accuracy : 0.5556
                 95% CI : (0.3076, 0.7847)
    No Information Rate : 0.5556
    P-Value [Acc > NIR] : 0.5966

                  Kappa : 0.0769

 Mcnemar's Test P-Value : 0.7237

            Sensitivity : 0.7000
            Specificity : 0.3750
         Pos Pred Value : 0.5833
         Neg Pred Value : 0.5000
             Prevalence : 0.5556
         Detection Rate : 0.3889
   Detection Prevalence : 0.6667
      Balanced Accuracy : 0.5375

       'Positive' Class : 1
```

In fact, this data does not tell us much about the true outcome of the game at all. Positive and negative predictive values are low, and we might be better off just flipping a coin. While we can't create any concrete reasons why solely from this data, we can conclude that this model does not consistently provide an accurate prediction.

Conclusions

In my own humble opinion, this game is the best rivalry in sports. Every piece of data can be thrown out the window when it comes to these two schools, and instead of trying to jump to

any conclusions, we're better off just going along for the ride. My full article breaking down these findings in a more compact, thorough way can be found on dukechronicle.com.