* Points vs possessions
* Condition on previous throw
  + Include instances where throw was not completed
* Before:
  + Looking at score % at start and end, and probability of completing that pass
  + Now looking at score % at start and end, including conditioning of when that pass is completed

# Research Question:

## Does conditioning on the location of the previous throw lead to better estimates for the score probability?

# Procedure:

## Parse data into possessions, including outcomes and throw sequences

## Create list of tuples including:

## Throw origin

## Throw destination

## Throw outcome

## Possession outcome

## Separate into training and testing data – perform calculations on training data

## Calculate unconditioned score probabilities

## Create matrix of locations

## Use k\_0 nearest neighbors to each location and average score rate

## Calculate conditioned score probabilities

## Create 4-d array (x\_0,y\_0,x\_1,y\_1)

## Use k\_1 nearest neighbors to each 4-d point and average score rate

## Compare unconditioned vs. conditioned score probabilities

## Estimate expected scoring probabilities of each testing data point (x\_0,y\_0,x\_1,y\_1) based on the trained model. For the unconditioned model limit the input to (x\_0,y\_0). Use knn with k=1 to select row of model to use.

## Calculate empirical scoring probabilities in testing data using appropriate knn method

## Calculate difference in expected and empirical for both conditioned and unconditioned sets. For the conditioned model calculate the difference in (x\_1,y\_1) arrays and take the average. This is the error for each model at that point

## Subtract the unconditioned model error from the conditioned model error at each point

## We now have a set of values that under H\_0 are distributed per *binomial(n,p=.5)*. We can thus calculate a p-value for H\_1, that p>.5

# Strategic Applications:

# Visualizations: