Section 1: Data Exploration

1) The swarm is more accurate against the spread for regular season games compared to playoff games; in regular season games the swarm has a 48.32% accuracy rate while in playoff games the accuracy rate is 50%. Upon performing a chi square test of independence we get a large p-value (.8606). This means we do not have evidence that there is an association between game type and the swarm correctly picking the outcome of the game against the spread. So, the difference in accuracy rate is not statistically significant.

2) The swarm was accurate in the month of July with a 62.5% accuracy rate. This rate is about 6% greater than the next most successful month.

3) (Classifying the strongest swarm pick each session by the pick with the greatest brainpower)
The average accuracy ATS for the strongest swarm pick each session is .5484; this means that the swarm has a 54.84% accuracy rate for the strongest pick in each session

4) (Classifying the weakest swarm picks as pick with the smallest brainpower)
The average accuracy ATS for the weakest favorite pick each session is .4111; this means that the swarm has a 41.11% accuracy rate for weakest favorite pick in each session.

Section 2: Model Building

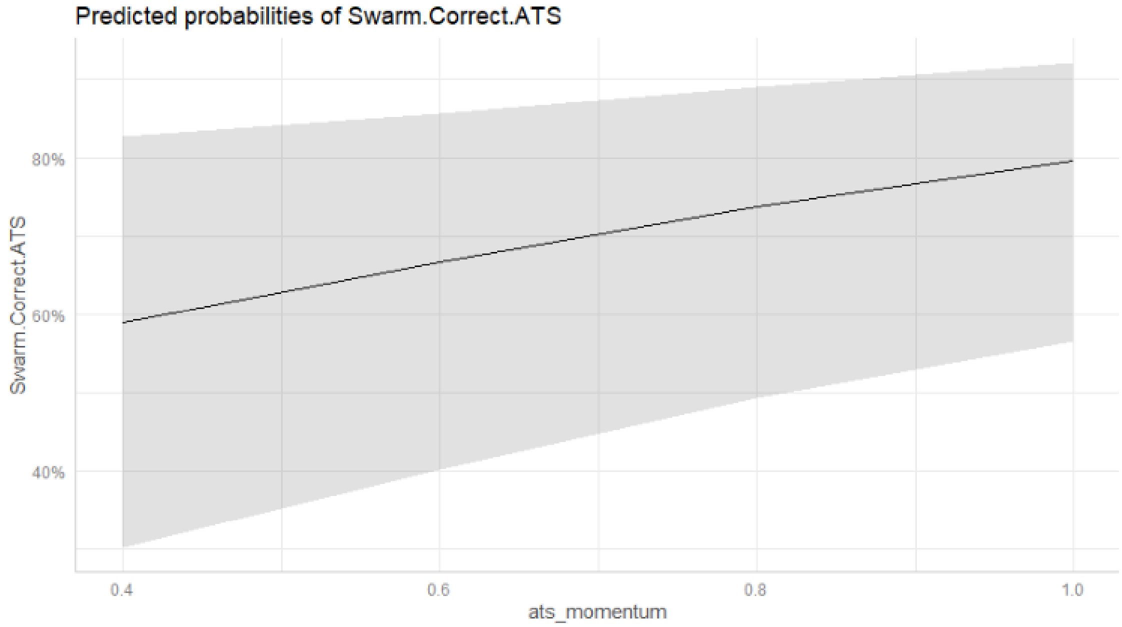
1a) I used a logistic regression model for predicting the probability of the swarm being correct for a given game.

1b) A logistic regression model is perfect for this type of data because this type of model predicts probabilities for a categorical binary response, which is exactly the case for this NBA dataset. The response is whether the swarm correctly picked the winner of the game against the spread which is binary and categorical (yes or no). The predictors implemented in my model are ats_momentum, Session.Num, and New.date (month in which the prediction took place). A good thing about this model is that it is small and easy to interpret with there being 3 main effects and no interactions. The variable that carries the most weight in the model is ats_momentum because it has the largest coefficient (value of 1.671). This is a good feature of the model because it was determined that ats_momentum was the most significant predictor of whether the swarm was correct or not.

2) The last session appears to be extremely accurate with an accuracy rate of over 90%. If I had 3 games to bet on I would choose the 139th, 183rd, and 187th games. This is because these games are in above the 95th percentile of ats_momentum values and are in the month of May which is the only month with a positive coefficient from the model which means when it is May the probability of the swarm being correct is higher.

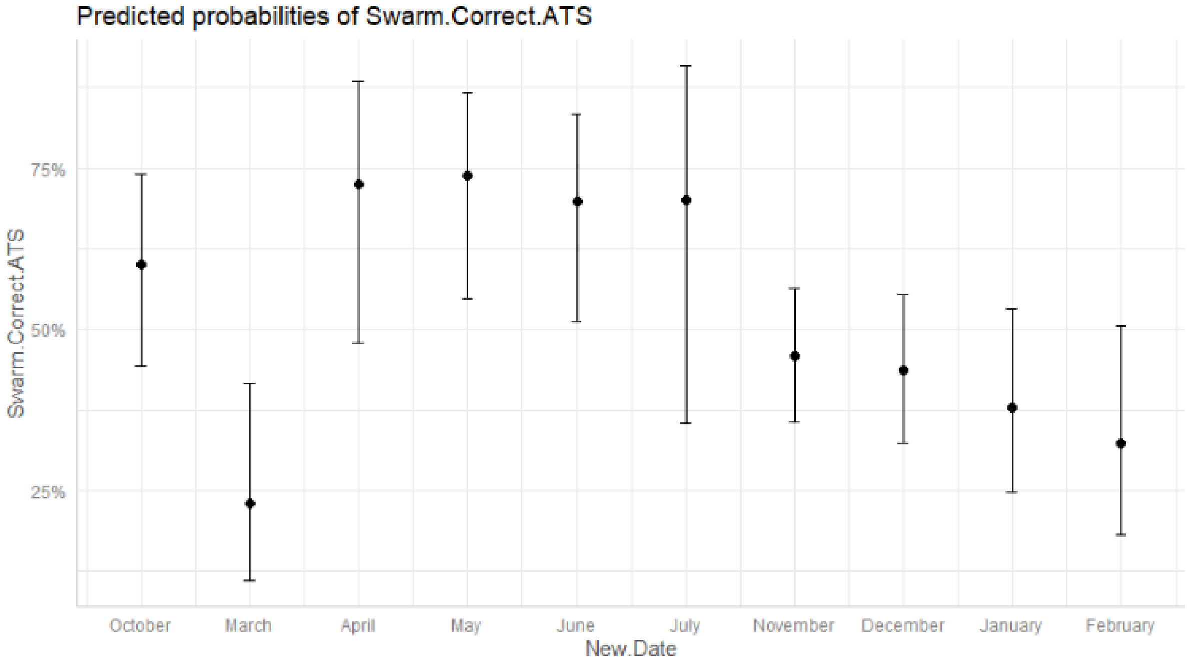
Section 3: Visual Representation

Graph 1: Probability of correctly picking ATS in regards to ats_momentum



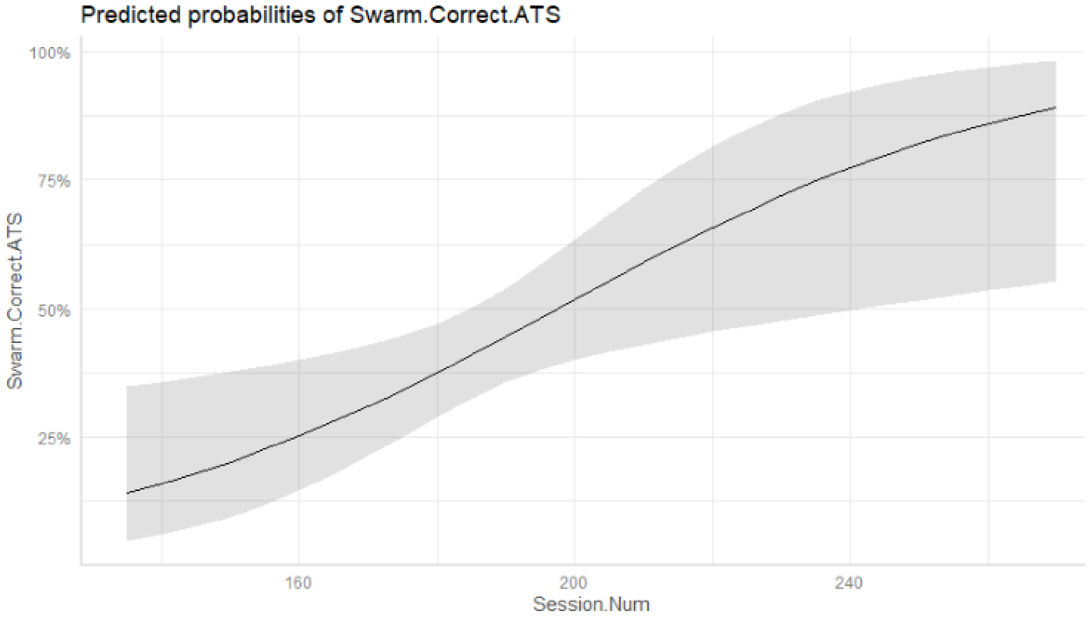
From this graph it is apparent that as ats_momentum increases, the probability of the swarm correctly picking the game against the spread increases as well. This is important because as stated before ats_momentum wad the strongest predictor of correctly picking against the spread so seeing that this variable affects the probability of correctly picking is a good thing.

Graph 2: Probability of correctly picking ATS in regards to month



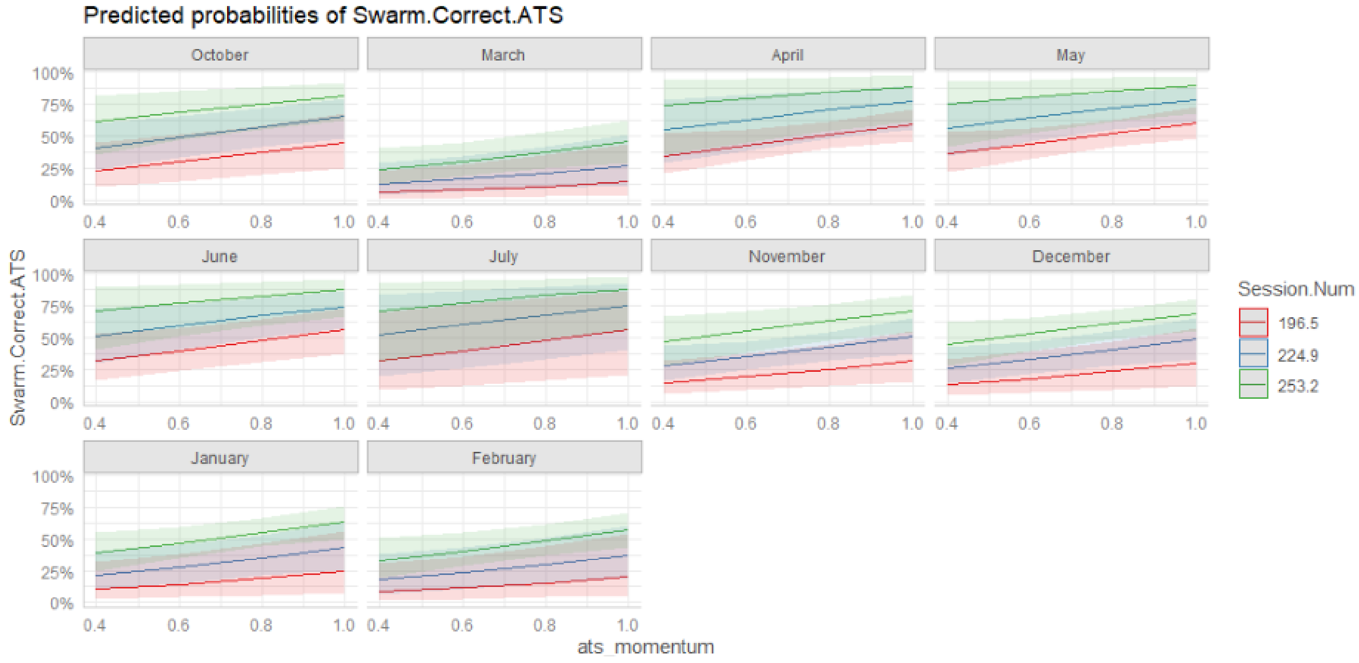
From this graph we see that the month in which the predictions take place can affect the probability of correctly picking the game. This is good to have in our model because s described in section 1, we found that months had varying accuracy rates.

Graph 3: Probability of correctly picking ATS in regards to session number



From this graph we see that the session number has an affect on the probability of correctly picking games ATS.

Graph 4: Logistic regression curves in regards to momentum grouped by month and session


Predicted probabilities of Swarm.Correct.ATS
October
March
April
May
June
July
November
December
January
February
Swarm.Correct.ATS
ats_momentum
Session.Num