| Problem Chosen | 2020 MCM/ICM Summary Sheet | Team Control Number |
|---|---|---|
| Problem D | | 2021087 |

## Executive Summary

We found qualitative indicators of team success for the Huskies soccer team. Success was measured in terms of the match outcome: win, loss, or tie. These indicators of success included the total number of passes per game, the type and frequency of node networks used, and the length of passes. The teams that had more passes throughout the match were more likely to win the match. These passes in a typical winning game included three different types of passing networks, two-node, three-node, and four-node, where the number of nodes indicates the players involved within the network. We found that while all three of these network types contributed to winning, the best indicator of success was an increase in the frequency of the three-node network utilization. We also found that shorter and simpler passes were more indicative of the team's success in the match. This makes sense, since if the distance between two teammates is shorter during a pass, the odds of the opposing team stealing the ball is reduced.

We created five quantitative models that focused mainly on whether it was a home or away game, the starting formation, and the starting player line-up. While these models did not have significant predictive power for the success of the team, it is important to note that we were not searching for estimations for model parameters. Our goal was to find different algorithms or processes that would try and classify a selected outcome, a Non Parametric classification. With this goal in mind, our qualitative determinations of success discussed above contributed to significant findings. In short, predicting the outcome of soccer games is much more complex than our models have the power to predict.

We hypothesized that other indicators of the team's success would include which individuals started each game. We determined the top three players in all three positions, defense, mid-field, and offense, based upon the frequency in which they received a passed ball throughout the season. Using these players as our starting line-up did not result in any significant increase in the odds of winning. However, our recommendation for designing a more effective team for the following season does urge the coaches to consider these players with the highest pass reception in forming their starting line-up.

Designing a successful team requires not only data analysis of what field events, such as passes, lead to a successful outcome, but also an understanding of the teammates' interactions. For a team to experience continual growth and lasting success, effective leadership and good communication are essential.

# Contents

# 1 Introduction

This problem focuses on teaming strategies. Specifically, on how the assembly of teams optimize efficiency and performance. Factors such as the interactions between teammates, the performance capabilities of each individual, the type and style of leadership, and how to best combine individual talents for achieving a specific goal. To analyze these essential factors in creating an efficient and high performing team, a soccer team's, the Huskies', data was collected and this data was analyzed.

## 1.1 Problem Overview

The goal for analyzing the Huskies' team dynamics was first to form an understanding of how they operate. We were given three different data sets detailing different aspects of the last 38 soccer games against 19 opponents, 2 games per opponent. Some of the key aspects recorded in the data sets included the types of events within the games, duels, passes, goal kicks, etc., and the $x$ and $y$ coordinates of each player's position along with the coordinates of the ball after each event. Analyzing this data helped us to acquire an understanding of which players are pivotal to team success, which players are frequently in play, what does a win compared to a tie or loss look like in terms of certain type of event occurrences, and so on.

The second step was to determine which components and team interactions are the most important for the overall success of the team in terms of winning. After identifying these essential teammate interactions and individual components, we created strategies to increase the likelihood of victory for the following season. Specifically we created:

- a network for the ball passing between players to identify team formations
- performance indicators which reflect the success of teamwork
- teamwork models which gave us insight into successful structural strategies and allowed us to form recommendations for an improved subsequent season.
- a generalized understanding of the essential components for any team to achieve success, and what other aspects of teamwork still need to be captured.

## 1.2 Background Research

The article on network motifs in football, by GÜRSAKAL, N and others, analyzes the critical points in a football game. These points are the "driver nodes" and they can be used to control the network which is built up of motifs, or "small local connection patterns". In our models, we are using several different techniques to determine our own driving nodes. While this motif- driven network analysis paralleled our own thought process of how to initially tackle the problem, this method was unable to account for the positions of the players in the soccer field. However, this source supported the drive behind our network analysis, shown in both our initial explorations and in the formation of our models.

## 1.3 Assumptions

1. The field and weather are in ideal conditions.

2. The best indicator of successful teamwork is the match outcome: a win, loss, or tie.

3. The player ID number signifies the same player throughout the entire season.

4. The players who are passed the ball more frequently are more important players than those who are passed the ball less frequently. The reasoning being that the player is skilled in reading the field and anticipating the best position for receiving the soccer ball. Additionally, teammates know which players are the strongest and most likely to achieve success, therefore, they are at an increased likelihood of passing the ball to these players.

5. The players who played this season will either continue into next soccer season, or they will be replaced in a 1:1 ratio by players with equivalent soccer skills.

6. Players whose location is typically near the pathway of the ball will have more influence on the pathway of the ball than players who are typically positioned further away from the ball's pathway.

7. The network used for passing the ball during successful games indicates a desirable ball trajectory pathway which is indicative of winning.

8. Players who are starting, beginning the match on the field, have a higher average time in play than players who are not starting.

# 2 Initial Exploration

In this initial exploration of the Huskies' team we created plots, calculated frequencies of players receiving the soccer ball, analyzed the typically time a player spent with the ball, and found the average location of a player on the soccer field. This exploration provided a rudimentary overview for what events or positions typically predicted winning the match.

## 2.1 Overview of Matches

This last season occurred over the course of 38 matches with 19 different opponents. Since each match occurred twice with the same opponent, a comparison could be established between the first and second match. This comparison was useful in that it should show the growth of the team over the course of the season. The first match is in early season and the second match is later in the season, this limits the variability introduced when comparing one match with one opponent with another match with a different opponent. Plots were created to compare the first and second match against each opponent. These plots allowed for the creation of some basic generalizations. In plotting all of the matches, we found that the two most common events were always duels and passes. Additionally, the likelihood of a win or loss seemed to correlate with the number of passes that each team has, see figure 8 and figure 9 in the Appendix.

Figure 1 and figure 2 illustrate the trend of passes influencing the match outcome. Figure 1 shows the two matches against Opponent 2. The first match-up, match 2, resulted in a tie. Although the opponent has more passes than the Huskies, they had an even split on the duels were fairly equal in all other event types. In the second match-up, match 32, the Huskies lost. A comparison between the two matches suggests that the loss in the second game was a result of the significant increase in the number of opponent passes as all other events remained fairly equal. Figure 2 seems to almost appear in contrast to the first graph. In the Huskies' first match against Opponent 10 they lost despite having a higher number of passes. In their second match, the Huskies won despite having less passes than their opponents. Rather than dismissing this plot as an outlier against the typical trend found throughout the matches, which bares a closer resemblance to figure 1, we recognized that this plot might indicate a more complicated system at play. Generally, an increase in passes correlates with a win. A rational reason behind this would be that passes indicate an increase in the team's time with the ball. This would increase in control over the ball would increase a team's likelihood of scoring goals as they are advancing the ball. However, where this assumption falls short, as is shown in figure 2, is an increased number of passes being due to pressure not strategy. The team may be increasing their passes because they are being pressed by the opposing team into making truncated plays until the opposing team regains control over the ball. Therefore, although an increase in passes typically correlates with a team winning, it cannot always be assumed to be so.


Figure 1: This plot shows the two matches against Opponent 2. The first game, match 2, was a tie, and the second game, match 32, was a loss for the Huskies. This graph represents the general relationship seen with an increase in a team's passes correlating with an increased likelihood of them winning

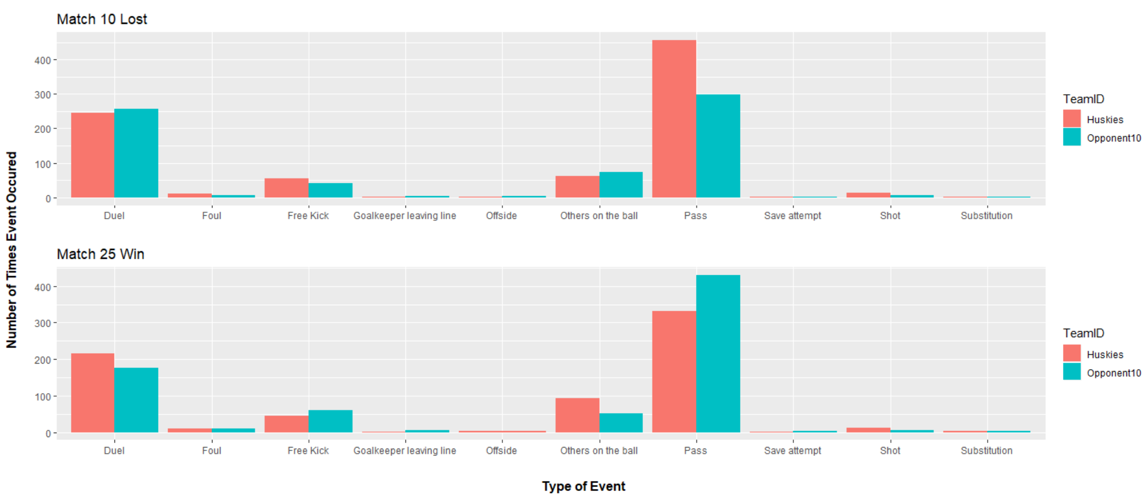


Figure 2: This plot shows the two matches against Opponent 10. The first game, match 10, was a loss for the Huskies, and the second game, match 25, was a win for them. This graph shows an atypical trend where a decrease in passes leads to a team winning.

## 2.2 Duration of Ball Possession

We decided to determine the total time, in seconds, each team had control over the ball during the course of a match. Our premise was that a direct relationship would exist between winning and an increased duration of ball possession. Logically, the longer a team has control over the ball, the higher the likelihood of scoring a goal. We used the “fullevent” data set to find the time for each event. This time, in seconds, included the team whose event it was. We used this information to determine the total amount of time that each team had the ball for. We did this by taking the difference in time between the team’s first event and the time at which the opposing team gained possession of the ball. We then added up the total time during which a team had the ball in play for each match. This was done for all 38 matches for both the Huskies and their opponents, Table 1. After comparing the Huskies’ ball possession duration and outcome from match to match, Appendix table 10, we did not immediately see the positive correlation we had anticipated. However, despite missing evidence of a concrete connection between possession duration and victory, we still decided to include the duration of ball possession as an indicator of success within our models.

| Match | Huskies(sec) | Opponent(sec) | Match | Huskies(sec) | Opponent(sec) |
|---|---|---|---|---|---|
| **1** | 4298.599776 | 7136.800522 | **20** | 6985.664872 | 4210.092741 |
| **2** | 4123.748872 | 4390.915829 | **21** | 4425.429521 | 7329.931547 |
| **3** | 6882.023299 | 4533.051236 | **22** | 4171.407494 | 7154.11345 |
| **4** | 7245.13435 | 4281.108875 | **23** | 4225.278088 | 4121.748638 |
| **5** | 4358.548142 | 4003.674468 | **24** | 7703.163876 | 4092.318841 |
| **6** | 4239.97362 | 7399.580241 | **25** | 4186.753148 | 7362.556666 |
| **7** | 4666.203711 | 6767.938298 | **26** | 7024.283492 | 4248.582845 |
| **8** | 4214.429242 | 7161.598513 | **27** | 6924.363076 | 4475.613972 |
| **9** | 7312.877853 | 4247.839916 | **28** | 4275.07011 | 6972.696085 |
| **10** | 4562.842066 | 6926.723602 | **29** | 4368.750231 | 7071.18985 |
| **11** | 4387.071935 | 4666.228898 | **30** | 4063.449535 | 7357.642705 |
| **12** | 7403.252566 | 4087.070822 | **31** | 4149.159067 | 4334.865165 |
| **13** | 4318.158241 | 7259.689136 | **32** | 3638.936869 | 7647.10658 |
| **14** | 4024.604468 | 4498.611022 | **33** | 4173.768142 | 4080.781404 |
| **15** | 7084.864953 | 4399.783334 | **34** | 4037.658704 | 7353.411907 |
| **16** | 3809.842479 | 7668.869899 | **35** | 4092.153184 | 7434.086322 |
| **17** | 7116.861827 | 4354.933116 | **36** | 6964.340523 | 4477.561581 |
| **18** | 4037.745455 | 4393.517466 | **37** | 4626.454664 | 3938.673289 |
| **19** | 6947.960628 | 4402.241424 | **38** | 4355.934946 | 4275.878972 |

Table 1: This table shows each team’s duration of ball possession in seconds.

## 2.3 Frequency of Player Possession

To determine the most influential players of the Huskies’ team, we made a few assumptions. Three positions exist in the soccer field: defense, mid-field, and forward. In the data that was given, we had the players who received a passed ball as well as the position and identification of that player. We assumed that the players who received more passes were more crucial to the winning of a match. This assumption was based on a two-fold reasoning. First, a player who can read the field and anticipate where to position him or herself to increase the odds of receiving a pass has more soccer prowess than a player who cannot. Second, a player’s frequent reception of a passed ball signifies that his or her teammates are passing the ball due to a belief in this athlete’s ability to successfully obtain possession of the passed ball and advance it towards the opponent’s goal.

Using this logic and the given data, we determined the frequency of each player’s pass reception. We found this frequency by taking the number of passes each Huskies player received throughout a match and dividing it by the total number of Huskies passes in the match. We used this method to determine the frequency of each player, in all three positions, for all 38 matches of the season. To determine the players who had the highest frequency of pass reception, we took the sum of each player's frequency over the season and divided by the number of games they played in. From here, we were able to determine the top three players in each position. Table 2 shows the top three players by position. Figure 3 shows the top defensive player's pass reception frequency by match. Each of the top player's pass reception over the course of the season was analyzed to determine if frequencies were increasing or decreasing. However, frequencies remained fairly constant match to match. After calculating each players frequencies we were able to calculate the teams overall pass success rate from looking into the amount of incomplete passes. We found that the Huskies had an average season passing success rate of 63.7%.

| | Name | Frequency(%) | Matches Played |
|---|---|---|---|
| Defense: | | | |
| 1) | Huskies_D4 | 7.26 | 23 |
| 2) | Huskies_D5 | 7.08 | 22 |
| 3) | Huskies_D3 | 6.76 | 24 |
| Mid-Field: | | | |
| 1) | Huskies_M1 | 7.99 | 33 |
| 2) | Huskies_M3 | 6.22 | 30 |
| 3) | Huskies_M6 | 5.36 | 27 |
| Forward: | | | |
| 1) | Huskies_F2 | 7.81 | 31 |
| 2) | Huskies_F6 | 5.64 | 14 |
| 3) | Huskies_F1 | 4.30 | 32 |

Table 2: This table shows the top three players of each position based off of their mean percentage frequencies calculating using the number of matches they played in.

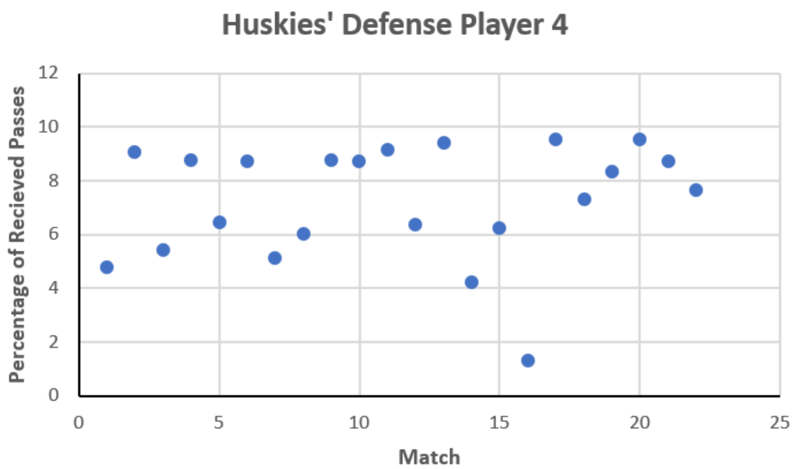


Figure 3: This is a plot of the top defense player, player number 4. This plot demonstrates that the pass reception frequencies, shown here in terms of percentages, remained fairly constant throughout the season.

## 2.4 Player's Average Position

We also formed the assumption that a player whose average position is closer to the ball's typical trajectory will have an increased likelihood of interacting with and controlling the ball's motion. This is opposed to a player whose average position is further away from the ball's traveling pathway. To determine where a player's average position was, we first took the given $x$ and $y$ coordinates for the origin of each event, since these coordinates signified the player's position at the start of an event. We then separated out the these locational coordinates by an individual basis, and took the average $x$ and $y$ coordinates for each player. Then to get the relative movement of each player, we took the standard deviation for each player's average position. This formed a domain of activity for each player.

We used the players' typical domains to determine the their formation on a match by match basis. We found six different formations present throughout the entirety of the matches: 4:5:1, 5:3:2, 4:3:3, 4:1:5, 5:4:1, and 4:4:2. The first number in each sequence is the defensive players', the second is the mid-field players', and the last is the forward players'.

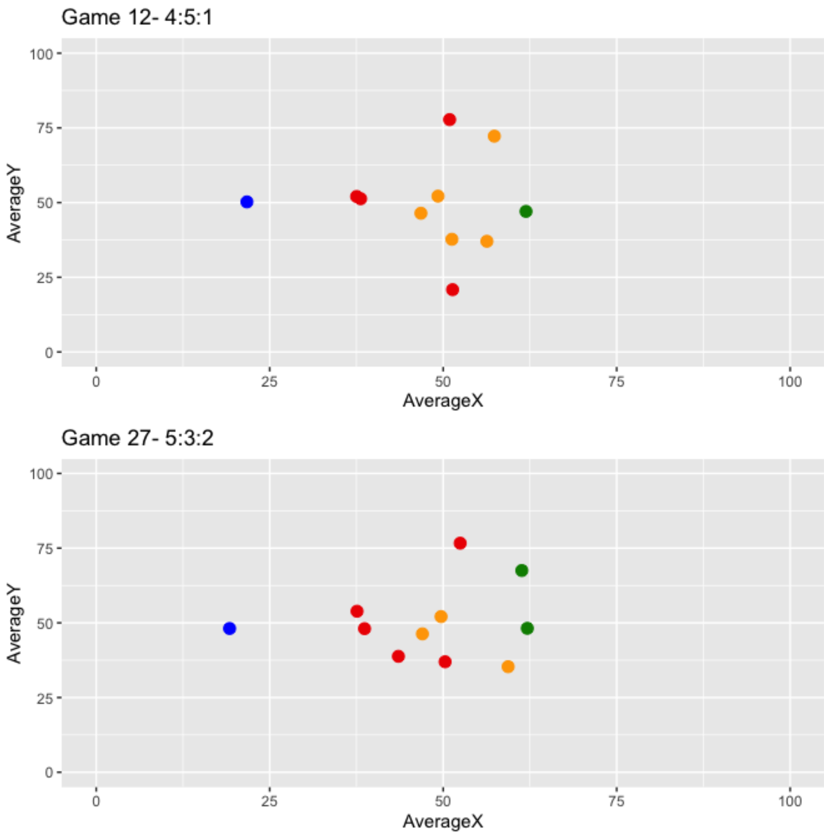


Figure 4: These formation plots show defense(red): mid-field(yellow): forward positions (green) with the specific formations 4:5:1 on top and 5:3:2 on bottom. The blue point signifies the goal keeper.

## 2.5 Most Frequented Passing Paths

We found the three most frequented types of paths, two-node, three-node and four-node, for each successful match, where success was measured in terms of winning. Two-node paths represent a single pass between two players. These pathways were found by looking at both the origin and the destination of a passing event. The Three-node paths represented a double pass between three players. These pathways were found by looking at the origin and the destination of a passing event, and then determining if the following passing event's origin matched the previous event's destination. If it did, then this event's destination was added to the pathway which resulted in the creation of our three-node pathways. Four-node paths were determined in the same manner as the three-node paths, but instead of only two passing events, three passing events were used. The three most frequented node paths for all three different pathway types are shown in table 3. It is important to note that these are the frequent pathways of only the Huskies' winning games, as we have assumed that successful pathways correlate with wins.

| | |
|---|---|
| **Two-Node** | $\text{Huskies}_F2 --> Huskies_M1$<br>$\text{Huskies}_M1 --> Huskies_F2$<br>$\text{Huskies}_D3 --> Huskies_F6$ |
| **Three-Node** | $\text{Huskies}_M1 --> Huskies_F2 --> Huskies_M1$<br>$\text{Huskies}_D3 --> Huskies_D1 --> Huskies_D4$<br>$\text{Huskies}_D1 --> Huskies_D2 --> Huskies_D3$ |
| **Four-Node** | $\text{Huskies}_D3 --> Huskies_G1 --> Huskies_F2 --> Huskies_D4$<br>$\text{Huskies}_D4 --> Huskies_F5 --> Huskies_D1 --> Huskies_D4$<br>$\text{Huskies}_D5 --> Huskies_F1 --> Huskies_F2 --> Huskies_M6$ |

Table 3: Frequented passing paths throughout the games which were won.

# 3 Teamwork Models

In our models, we have made the assumption that teamwork success is best measured by the outcome of the match, whether the team won, lost, or tied. For this reason, the following four non-parametric models were designed to try and predict the Huskies' match outcome using various combinations of predictors, and a random forest modeling technique.

## 3.1 Model One

Outcome as predicted by side - whether the team is at a home or away match, starting formation, and number of top players starting. This model resulted in a classification error rate of 60.53%. In table 4 below, we see the relative importance of each predictor, or how much it effects the model.

| **Predictor** | **Relative Importance** | **Rank** |
|---|---|---|
| Starting Formation | 47% | 1 |
| Side | 27% | 2 |
| Number of Top Players Starting | 26% | 3 |

Table 4: Relative Importance of each predictor.

## 3.2 Model Two

Outcome as predicted by both side and starting formation. This model resulted in a classification error of 55.26%. Below in table 5, we see the relative importance or how much each predictor effects the model.

| **Predictor** | **Relative Importance** | **Rank** |
|---|---|---|
| Starting Formation | 62% | 1 |
| Side | 38% | 2 |

Table 5: Relative Importance of each predictor.

### 3.3 Model Three

Outcome as predicted by both side and number of top players starting. This model resulted in a classification error of 47.37%. Below in table 7, we see the relative importance or how much each predictor effects the model.

| Predictor | Relative Importance | Rank |
|---|---|---|
| Number of Top Players Starting | 47.7% | 2 |
| Side | 52.3% | 1 |

Table 6: Relative Importance of each predictor.

### 3.4 Model Four

Outcome as predicted by only side. This model resulted in a classification error of 47.37%. Below in table 7, we see the relative importance or how much each predictor effects the model.

| Predictor | Relative Importance | Rank |
|---|---|---|
| Side | 100% | 1 |

Table 7: Relative Importance of each predictor.

### 3.5 Model Five

Outcome as predicted by only starting formation. This model resulted in a classification error of 57.89%. Below in table 8, we see the relative importance or how much each predictor effects the model.

| Predictor | Relative Importance | Rank |
|---|---|---|
| Starting Formation | 100% | 1 |

Table 8: Relative Importance of each predictor.

## 4 Results

When looking at the models above, we are able to tell from the classification error rate that each model leaves large room for improvement. Out of the five different models the best classification error rate of 47.37% which comes from Model Three and Model Four. Model Three is predicting a win, loss, or tie from whether or not the Huskies are home or away and also by the number of top players starting. We see from the relative importance, table 7, of Model 3, that the side is ranked as a better predictor than the number of top players starting. Model Four is predicting match outcome from side, again with an error rate of 47.37%. The error rate of 47.37% tells us that our model is comparable to a model that simple picks a coin. In other words, flipping a coin for the outcome would be just as effective as trying to predict the outcome from these variables. This is suggestive that predicting the outcome of a soccer games is much more complex than our models have the power to predict.

## 5 Sensitivity Analysis

In our sensitivity analysis, we sought to determine how a few of our key predictors for a team's success were influenced by changes.

### 5.1 Sensitivity of Top Players' Pass Receptions

Here we tested how sensitive the top players' pass reception frequency, shown in percents, was to both increases and decreases. The relevance behind checking this sensitivity was we needed to know how stable the selection of these top players was. For example, if the Huskies are relying on the top defensive player Huskies_D4 to maintain protection of the goal and this player has a rough day and is unable to successfully receive his or her typical number of pass receptions, what does the stability look-like. In other words, if a top player has a change in their typical pass reception frequency, does a new player replace this one in the top three position rankings. Table 9 shows the changes in percentage of pass reception frequency after a 10% increase and a 10% decrease. In some instances, when we decreased the frequency by 10% a new top player was selected to replace the poorly performing one. This new player who joins the top three ranking is specified in the far right column.

The sensitivity of these top players performance is important for the coaches to recognize in order to optimize the team's performance. If a player is sensitive to decreases in the frequency of pass receptions, then the coaches should know whom to replace this player with. This is essential as the team relies on these high performing players' abilities to read the field and communicate with their teammates in order to successfully receive a pass.

| | Name | Frequency(%) -10% | Frequency(%) | Frequency(%) +10% | New Top Player |
|---|---|---|---|---|---|
| Defense: | | | | | |
| 1) | Huskies_D4 | 6.54 | 7.26 | 7.99 | Huskies_D7 |
| 2) | Huskies_D5 | 6.37 | 7.08 | 7.78 | Huskies_D7 |
| 3) | Huskies_D3 | 6.09 | 6.76 | 7.44 | Huskies_D7 |
| Mid-Field: | | | | | |
| 1) | Huskies_M1 | 7.19 | 7.99 | 8.79 | None |
| 2) | Huskies_M3 | 5.60 | 6.22 | 6.85 | None |
| 3) | Huskies_M6 | 4.82 | 5.36 | 4.73 | None |
| Forward: | | | | | |
| 1) | Huskies_F2 | 7.03 | 7.81 | 8.60 | None |
| 2) | Huskies_F6 | 5.07 | 5.64 | 6.20 | None |
| 3) | Huskies_F1 | 3.90 | 4.30 | 4.73 | Huskies_F4 |

Table 9: This table shows the top three players of each position and a 10% change within the frequency of pass receptions. If a player is sensitive to the 10% decrease then their replacement is shown on the right.

### 5.2 Sensitivity of a Player's Average Position

Since the player's position is pivotal to ball possession, we needed to determine the likelihood that each player was within their predicted field position. This is important for formulating successful strategies since a strategy which includes the player's typical position relies on the assumption that the player will be near their typical $x$ and $y$ coordinate.

Figure 5 shows the ellipse encompassing each player's field position. These domains represent one standard deviation away from the player's predicted $x$ and $y$ coordinate, shown in the figure below has dashed color-coordinated lines. This means that 96% of the passing events for that player will fall within that ellipse. A larger ellipse correlates with a greater domain and a decreased ability to predict the exact positioning of a player. Coaches can utilize this knowledge of each players' different domain sizes in order to device a plan that can accommodate the differing degrees of variability in each of the players' locations.

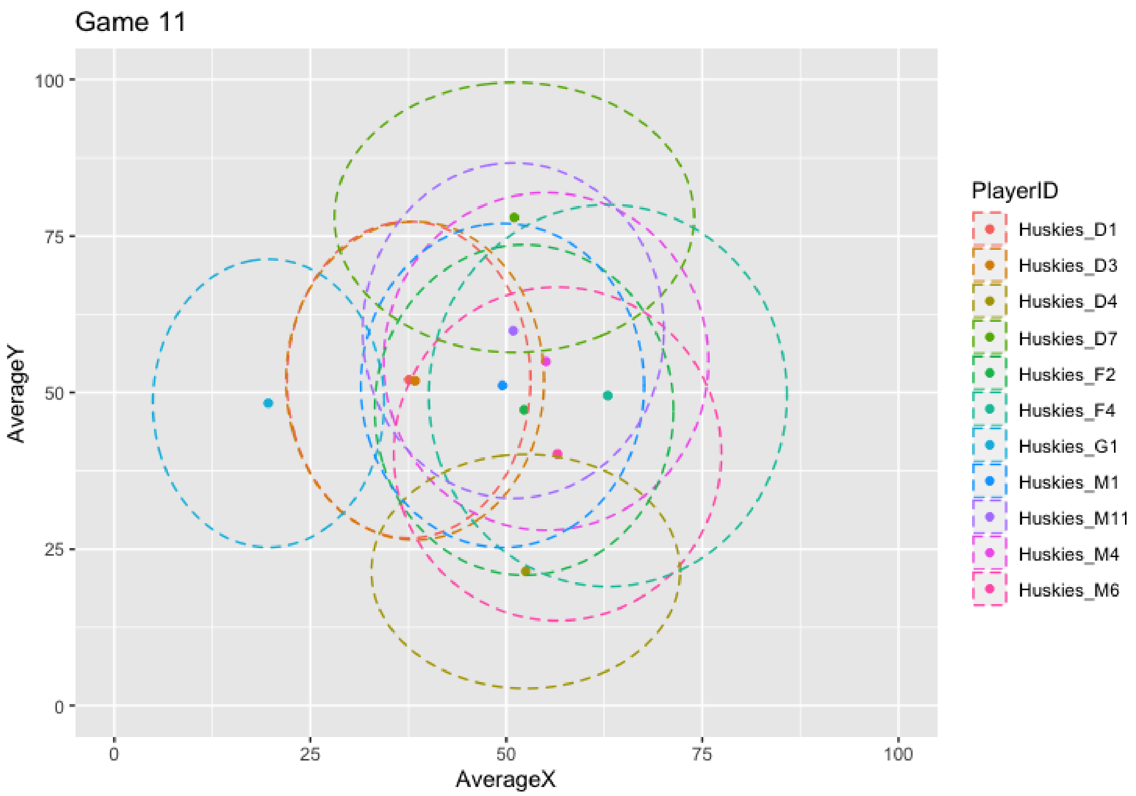


Figure 5: This plot shows the average position and standard deviation ellipse of each player during a 4:4:2 formation.

# 6 Generalized Results

When looking at the models above, we are able to tell from the classification error rate that each model leaves large room for improvement. In other words, any combination of side, player formation, and number of top players starting are not effective predictors of the outcome for a game.

However, as determined before, by finding the players that are most frequented, we can extrapolate which combinations of passes should be used to help the outcome for a game.

## 6.1 Designing a More Effective Team

Our suggestions for designing a more effective team for the Huskies next year is as follows.

1. While they should continue to use all of the same three passing node networks, they should increase the use of the three-node pathway. We determined that the three-node passing combinations are most effective for a successful game. In general, the games in which the Huskies did best were the ones during which three-node passing combinations were used with an increased frequency.

2. Shorter passes are simpler and less likely to result in loosing the ball to the opposing team. In the exploration of the matches, we saw a correlation between an increase in passing events, and an increase in wins.

3. The top three players in each of the team's three positions, determine by the frequency of receiving a passed ball, should be included in the game earlier. Since these players have been shown to have the

highest frequency of successful passes, and an increase of successful passes correlates with winning, getting these players into the game earlier will potentially increase the odds of winning the match.

4. The top players who are sensitive to a decrease in their pass reception should be exchanged for the suggested alternative player, shown in table 9, if they begin to struggle with receiving the passes.

### 6.2 Teamwork Aspects Still Missing

Many different factors influence a team's performance. In the Huskies' case, we were able to utilize the details given for each match, such as event types, duration of ball possession, etc., to create models for optimizing team performance. However, we were missing data on one of the most crucial aspects of any team, the interpersonal relationships. A team can only function efficiently if they are able to communicate and work together effectively. Although details given on the play by play of a match can give insight into the effectiveness of a team's ability to work together, it is only an indicator of camaraderie and effective communication. A data set which focuses on teammate interactions would greatly improve the predictive accuracy of models created to enhance team success. Ideally, this data set would include: the team captain or captains, the age of each athlete, the conditions in which the athletes live, the training style of the coaches, and the turnover of players. All of these aspects are difficult to quantify and record in a data set, however, they hold significant control over the effectiveness of a team.

In developing generalized models for any team's performance, it is important to remember that a successful team relies on supportive teammate interactions and great leadership. If we can find a way to effectively capture the human-human interaction necessary to create an effective team, then we can begin to develop models which account for the power of these interactions. These models will demonstrate success as the product of humanity's ability and desire to work in cooperation with other individuals to achieve arduous tasks which are difficult or impossible to achieve alone.

## 7 Conclusion

In conclusion, we found good qualitative predictors even though the quantitative models did not have predictive power. As stated above, the team would benefit from increasing the frequency of short and simple three node passes. Practicing these passing paths will increase the trust between players involved in the play and in turn increase the overall frequency of successful passes for both individual players and the team as a whole. This increase of successful passes would then result in better players and an increase of winning games.

### 7.1 Overall Strengths

We analyzed each of the team's matches as a whole. This holistic viewing of each match, and each victory's successful patterns, throughout the season helped us to maintain its integrity. This integrity might have been lost if the successful and unsuccessful matches were summed and dissected into individual components.

Additionally, we explored a variety of different factors to determine what elements may contribute to the success of the team. The factors analyzed in each match, and over the season, included: the number of passes, the time each team had control of the ball, the frequency each player received a pass, and so on. The analysis of this variety of factors gave us insight into the complex interactions of the team's successful or unsuccessful performance.

### 7.2 Overall Weakness

Soccer is a very complex sport and it is very difficult to be able to predict quantitative success from purely qualitative factors. Individual's ability to make decisions makes it difficult to be able to predict exactly how players are going to move and decide to pass the ball. Furthermore, it is then almost impossible to predict how opponents and other players are going to react to a given event. Lastly, the results discussed above, are extremely specific to the Huskies', and would be difficult to apply to different teams. This would be because each team using unique formations and has a unique set of players with various skills.

## 8 Appendix

### 8.1 Plots

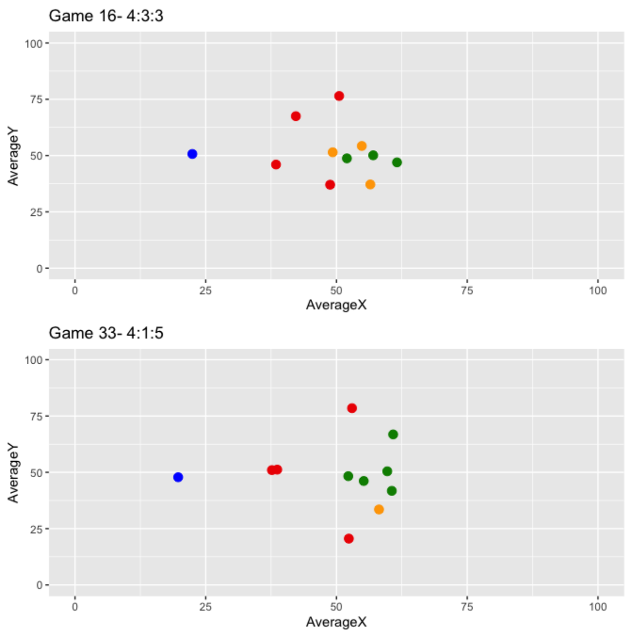


Figure 6: This plot shows the 4:3:3 and 4:1:5 formations.

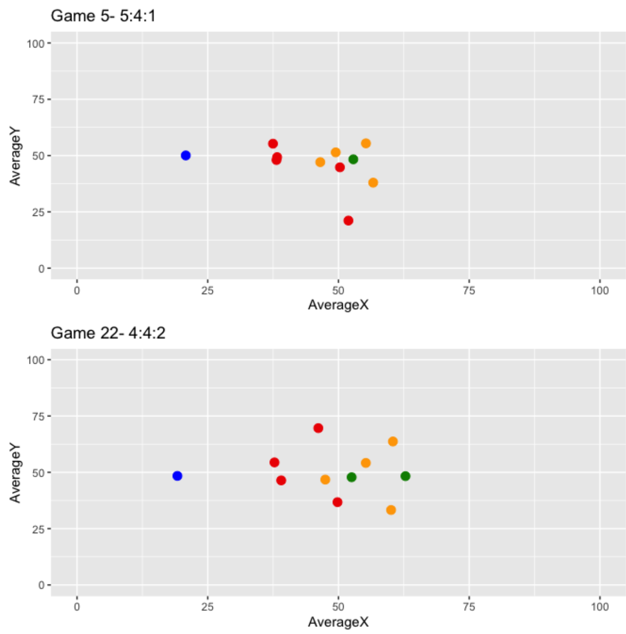


Figure 7: This plot shows the 5:4:1 and 4:4:2 formations.

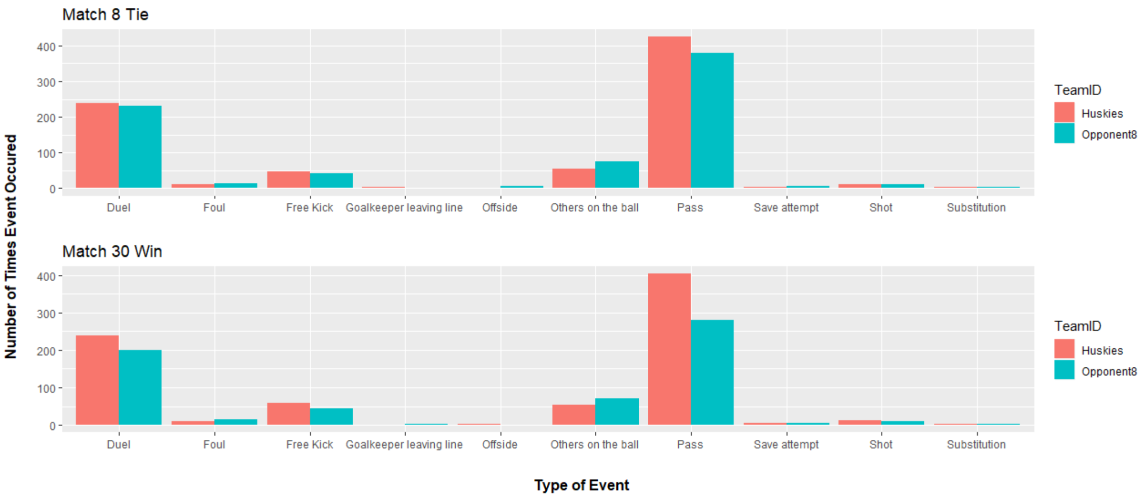


Figure 8: An increased numbers of passes generally correlates with winning.

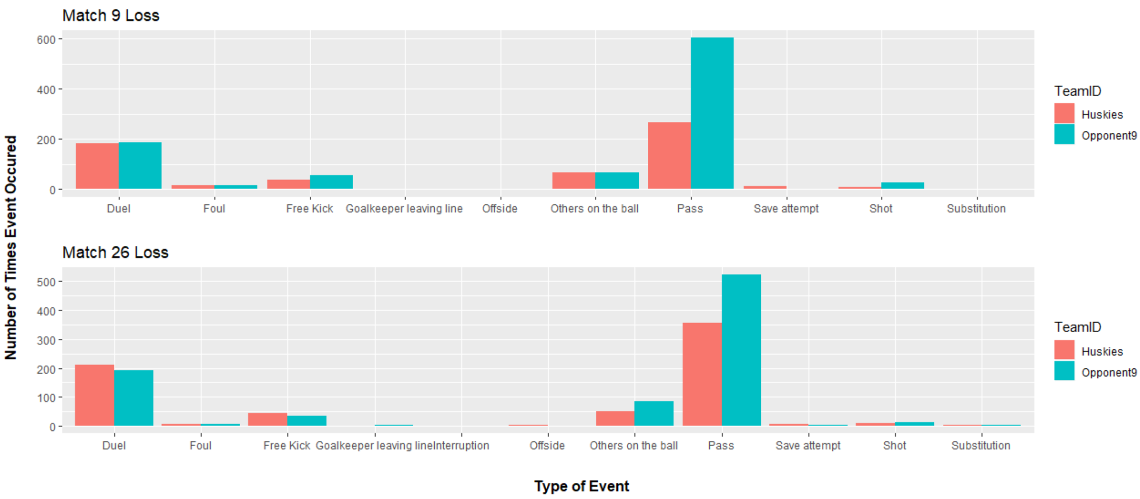


Figure 9: An increased numbers of passes generally correlates with winning.

## 8.2 Tables

| Match | Result | Match | Result |
|---|---|---|---|
| **1** | win | **20** | tie |
| **2** | tie | **21** | loss |
| **3** | loss | **22** | loss |
| **4** | loss | **23** | loss |
| **5** | loss | **24** | tie |
| **6** | win | **25** | win |
| **7** | loss | **26** | loss |
| **8** | tie | **27** | win |
| **9** | loss | **28** | loss |
| **10** | loss | **29** | loss |
| **11** | win | **30** | win |
| **12** | tie | **31** | win |
| **13** | loss | **32** | loss |
| **14** | win | **33** | tie |
| **15** | win | **34** | tie |
| **16** | tie | **35** | win |
| **17** | win | **36** | win |
| **18** | win | **37** | tie |
| **19** | tie | **38** | loss |

Table 10: Shows the Huskies wins, ties, and losses for each match, obtained from the given “matches” data set

# 9 References


1. Adjacency matrix. https://en.wikipedia.org/wiki/Adjacency_matrix

2. Buldú, J.M., Busquets, J., Echegoyen, I. et al. (2019). Defining a historic football team: Using Network Science to analyze Guardiola’s F.C. Barcelona. Sci Rep, 9, 13602.

3. Cintia, P., Giannotti, F., Pappalardo, L., Pedreschi, D., Malvaldi, M. (2015). The harsh rule of the goals: Data-driven performance indicators for football teams. 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA), 1-10, 7344823.

4. Duch J., Waitzman J.S., Amaral L.A.N. (2010). Quantifying the performance of individual players in a team activity. PLoS ONE, 5: e10937.

5. GÜRSAKAL, N., YILMAZ, F., ÇOBANOĞLU, H., ÇAĞLIYOR, S. (2018). Network Motifs in Football. Turkish Journal of Sport and Exercise, 20 (3), 263-272.

6. Gyarmati L, Hefeeda M. Analyzing in-game movements of soccer players at scale. Qatar Computing Research Institute, HBKU. 2016.

7. Sadler J, Introduction to Network Analysis with R. 2017. https://www.jessesadler.com/post/network-analysis-with-r/

8. Soccer Movement Positional Running. 2019. https://soccer-training-info.com/soccer_positional_running/