A deep dive into what the statistics tell us about how the Slay-Ups basketball team performs; the good, the bad and the downright awful.

Slay-Ups

Informal Statistical Analysis Report

Archer Linning

Contents

1.0 Introduction

This report uses the most up-to-date statistics from the Slay-Ups two seasons, including grading, from their time within the Hibiscus competition. Whilst a majority win team with a win percentage of 85.71%, Slay-Ups have significant room for improvement. Hence, it is important to understand the areas by which they are lacking, or more likely, the areas that they have no hope of improving on but just so we know how shit they are.

Data from 28 games (however this number is constantly increasing) was utilised with 4 games missing play-by-play data. Using this data, it was possible to draw conclusions about Slay-Ups performance based upon wildly outrageous and just straight untrue assumptions. This report investigates the impact of offense, defence and fouls on winning as well as dives into the performance of the team, individuals, and lineups by minute.

2.0 Data Summary, Assumptions and Observations
2.1 Data Summary

The full span of data used for this investigation can be found online at:

https://github.com/ArcherLinning/Slay_Ups_Stats

To locate the data, simply download the CSV files within the "Data" folder.

A summary of the data collected can be seen below in table 1.

Variable Name	Meaning	Description
File Name: BBALL_STATS.csv		
game	Game number	This was done in a strange order. The first 5 games are the first 5 regular season games of the most recent season. The next 3 games are the grading games of the most recent season. Then games 9-27 are in order (grading first) of the previous season. Then game 28 (and onwards) are the most recent games.
for	Points scored for Slay-Ups	Self-explanatory
qtr1_for	Points scored for Slay-Ups in 1st quarter	
qtr2_for	Points scored for Slay-Ups in 2nd quarter	
qtr3_for	Points scored for Slay-Ups in 3rd quarter	
qtr4_for	Points scored for Slay-Ups in 4th quarter	
against	Points scored against Slay-Ups	
qtr1_against	Points scored against Slay-Ups in 1st quarter	
qtr2_against	Points scored against Slay-Ups in 2nd quarter	
qtr3_against	Points scored against Slay-Ups in 3rd quarter	
qtr4_against	Points scored against Slay-Ups in 4th quarter	
fouls	Total fouls committed by Slay-Ups	
points_diff	Points differential at end of game	Slay-ups scored minus oppositions score. If positive, Slay-ups won, if negative, Slay-Ups lost.
result	Win or Loss	Self-explanatory
season	The season the game was played	Format: YEAR/SEASON
opp	Opposition	Self-explanatory
File Name: Slay-Ups-By_Minute.csv		
minute	Minute of the game	Each minute refers to the previous 60 seconds. E.g. 3rd minute is from time 2:01-3:00.
game#_off	Points scored for Slay-Ups in game number #	Self-explanatory
game#_def	Points scored against Slay-Ups in game number #	

Table 1: Data summary.

2.2 Assumptions and Observations

There were various assumptions that had to be made to yield any meaning from the data analysed. These

assumptions could be summarised in table 2.

Assumption	Effect	Likelihood of Assumption (Observation)
All data was correctly recorded by the scorers as well as accurately inputted by Archer.	Allows for accurate results.	*Somewhat likely* – some scorers are dumbasses who can't correctly input scores, and Archer is only human so could've made minimal errors in data processing, however most of the data seems to have been recorded accurately.
All substitutions occur at the exact minute	Allows to subset data to separate lineup and player performance.	*Impossible* – this never happens, so this is obviously a massive assumption that almost completely invalidates all individual and lineup statistics.
All players played all games	Allows to subset data to separate lineup and player performance.	*Impossible* – we know this is untrue, Dan's missed a bunch of games from work, Chapman missed a couple from injury, Mitch & Anton missed one or two, we've had fill-ins, etc. Again, pretty much invalidates all individual and lineup statistics.
All offensive a defensive output in game is accredited to the entire lineup	Allows to analyse lineup and player performances.	*Highly unlikely* – it doesn't take a genius to figure out that we score more points on average when Joseph and James are on the court, so the more playing time other players get with them, the better their offensive rating is. Similarly, if Archer, Dan and/or Anton are on the court, it is somewhat obvious that there is going to be less points scored against as we have greater size resulting in less chance of conceding offensive rebounds for second chance points, as well as having a greater defensive presence within the paint due to size. Hence, other players who are on the court during these times will benefit defensively.
All games play 40mins	Allows for consistency	*Impossible* – we had one game go to overtime and nearly every game loses time due to timeouts/late starts, etc.

Table 2: Assumptions & Observations

# 3.0 Analysis and Visualization

3.1 Influence on Winning

Firstly, the overall team statistics were investigated to understand what influences winning the most for Slay-Ups. Due to such a high win percentage, it was evident that the games were Slay-Ups lost would have the greatest weighting on this, however various trends could be derived from the data. The data summarised in this section was from "BBALL_STATS.csv". Summary statistics can be found in table 3.

Offense					
	for.	qtr1_for	qtr2_for	qtr3_for	qtr4_for
	Min. :38.00	Min. : 3.00	Min. : 6.00	Min. : 6.00	Min. : 4.00
	1st Qu.:40.75	1st Qu.: 9.75	1st Qu.: 9.00	1st Qu.: 9.75	1st Qu.:10.75
	Median :51.00	Median :11.50	Median :12.00	Median :12.00	Median :15.00
	Mean :54.29	Mean :12.21	Mean :13.46	Mean :13.04	Mean :16.33
	3rd Qu.:59.50	3rd Qu.:15.00	3rd Qu.:16.00	3rd Qu.:16.00	3rd Qu.:19.75
	Max. :94.00	Max. :23.00	Max. :29.00	Max. :29.00	Max. :36.00
		NA's :4	NA's :4	NA's :4	NA's :4
Defence	against	qtr1_against	qtr2_against	qtr3_against	qtr4_against
	Min. :26.00	Min. : 4.000	Min. : 0.000	Min. : 4.00	Min. : 7.00
	1st Qu.:30.75	1st Qu.: 7.000	1st Qu.: 5.000	1st Qu.: 8.00	1st Qu.: 7.75
	Median :37.50	Median : 8.000	Median : 8.000	Median :11.00	Median :10.00
	Mean :38.21	Mean : 8.458	Mean : 8.625	Mean :10.88	Mean :10.96
	3rd Qu.:46.00	3rd Qu.: 9.250	3rd Qu.:12.000	3rd Qu.:12.75	3rd Qu.:14.25
	Max. :60.00	Max. :16.000	Max. :21.000	Max. :20.00	Max. :18.00
		NA's :4	NA's :4	NA's :4	NA's :4
Other	fouls	points_diff			
	Min. : 4.000	Min. :-13.00			
	1st Qu.: 7.000	1st Qu.: 4.50			
	Median : 8.000	Median : 10.50			
	Mean : 9.417	Mean : 16.07			
	3rd Qu.:12.000	3rd Qu.: 28.75			
	Max. :16.000	Max. : 57.00			
	NA's :4				

Table 3: Summary statistics of "BBALL_STATS.csv".

3.1.1 Offense

Initially, we investigated the offensive output, i.e. points scored.
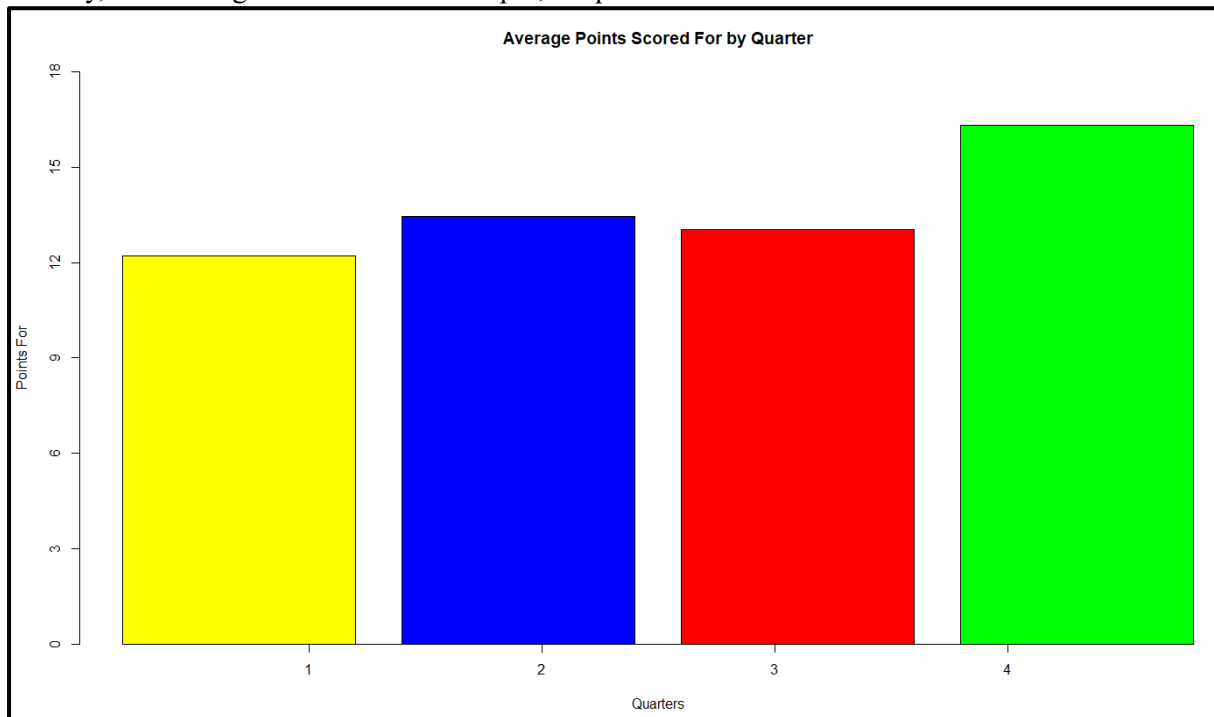
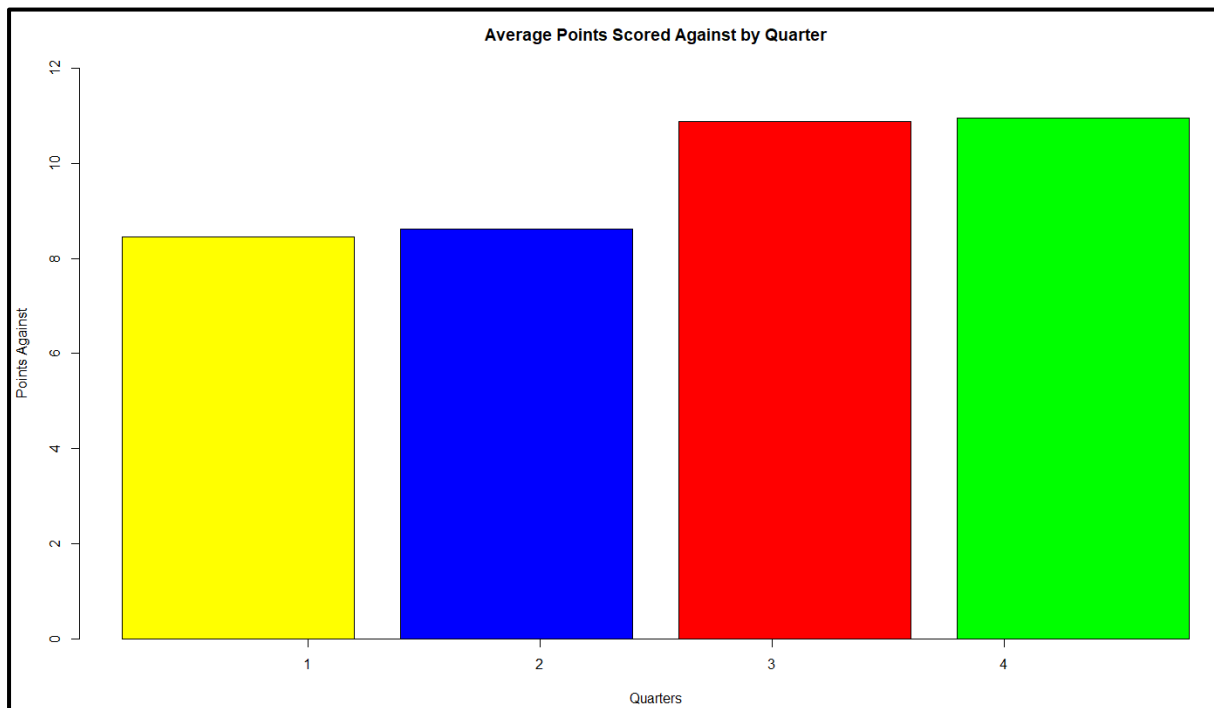*Graph 1: Average points scored for by quarter.*

Graph 1 illustrates that our slowest offensive quarter on average is the first quarter, followed by the third, then second and our best offensive output comes in the fourth quarter.

This is somewhat predictable as the first quarter usually loses time at the start. Additionally, players take time to "warm-up" and thus most of the quarter is spent warming into the game. Finally, the first quarter is the only period of the game where our two most offensively effective players are both on the bench (see section 3.3 & 3.4).

Furthermore, the fourth quarter has the greatest offensive output as usually defences become tired and hence it is easier to create shots which leads to more points. Additionally, the fourth quarter includes what is known as "garbage time" – the last couple of minutes in a game where both teams know the game is over as the teams a separated by too many points. During this time defences usually give up, allowing the offensive team to score easy points.

3.1.2 Defence

Next, we investigated Slay-Ups' defence i.e. points scored against.
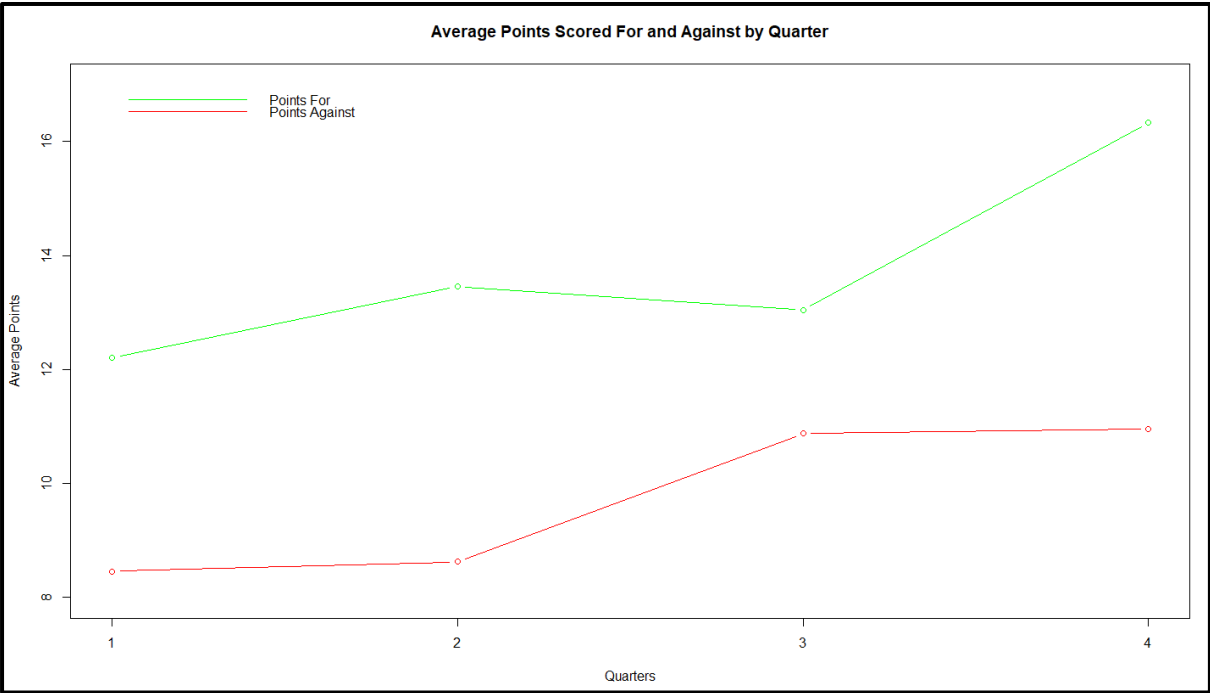

Average Points Scored Against by Quarter

*Graph 2: Average points scored against by quarter.*

From graph 2, Slay-Ups concede the most points during the second half, with the fourth quarter conceding slightly more points than the third on average (10.96 vs. 10.88). The next most points conceded is in the second which is closely followed by the first. Like the offense, the first quarter defence would most likely be attributed to the opposition's offense taking time to "warm-up" as well as lost time due to a delayed start. Interestingly however, the fourth quarter offensive trend is not as strongly reflected within the defensive data, with the third quarter effectively conceding the same number of points on average even though the fourth quarter includes the aforementioned "garbage time" as well as hosting the players at their most weary. However, the lacklustre defence seen in the third quarter could be due to the lack of defensive presence and size on the court during this period. Within the third quarter, 80% of the time there is two of Archer, Dan, Anton, or Joseph off – the 4 tallest players in the team. This lack of size could yield to greater chance of mismatches and hence could be the reason why the average points conceded is similar to that of the fourth quarter.
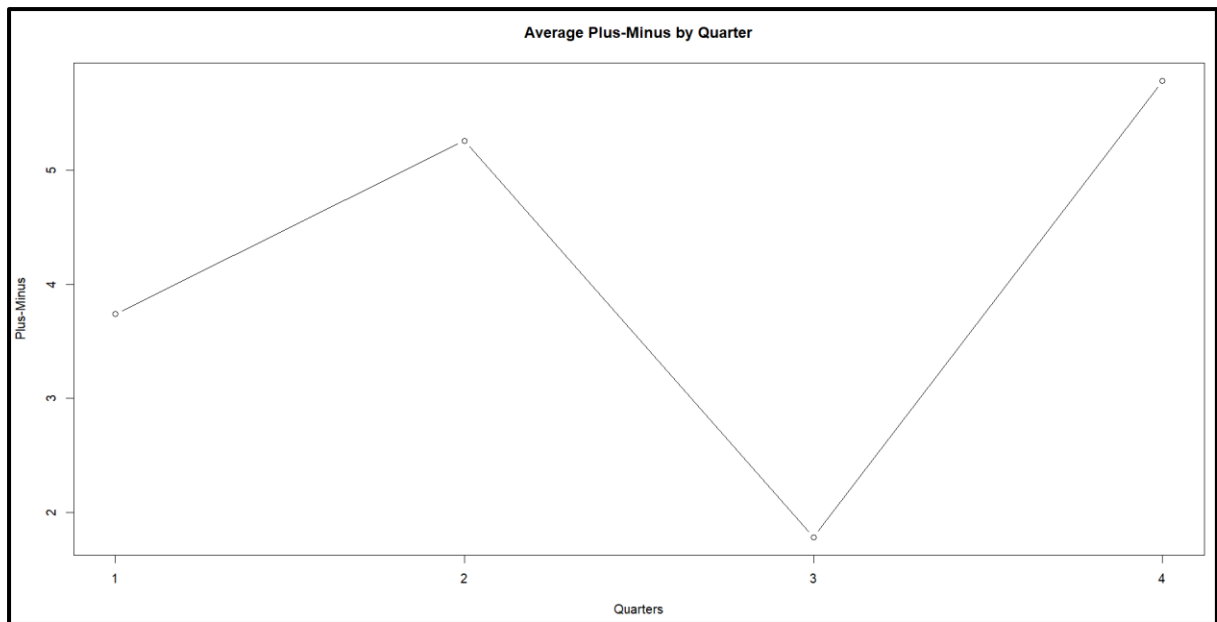
3.1.3   Plus-Minus

By utilising the offensive and defensive data, we can construct the plus minus by quarter to assess overall two-way performance of Slay-Ups.

*Graph 3: Average points scored for and against by quarter.*
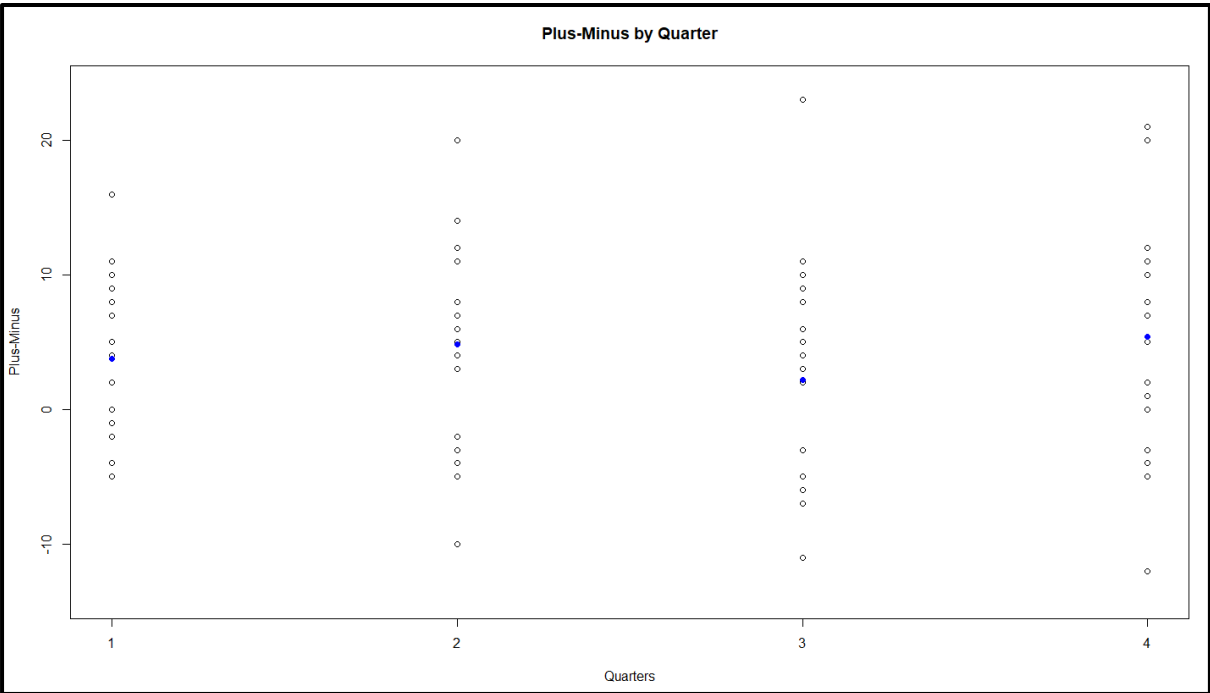
*Graph 4: Plus-minus range by quarter (blue is average).*

Graph 5: Average plus-minus by quarter.

3.1.4 Fouls

Another factor that could be influential on winning is the number of fouls given away. It is a standard trait of the Slay-Ups team to give away many fouls when frustrated as we have no self-control. These fouls range from pretty slaps to monster tackles and sometimes simply swearing at the bench/refs. These frustrations normally occur due to poor referee calls or simply because we are sore losers. Hence an investigation into the relationship between number of fouls committed and winning was analysed.
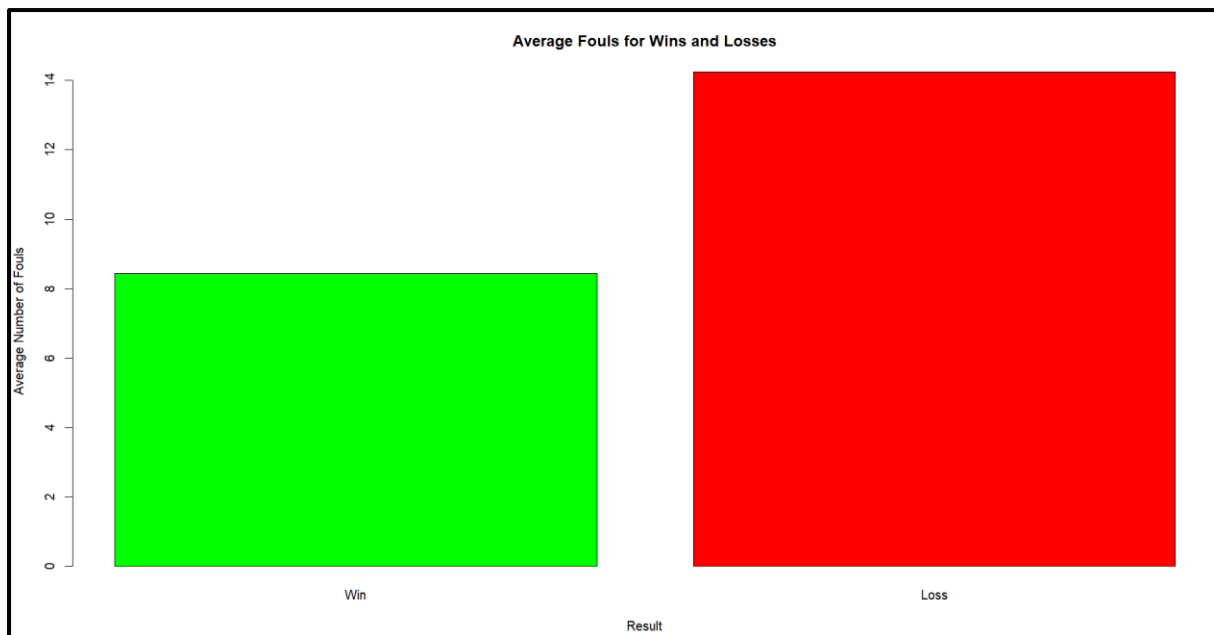
Graph 6: Average fouls for wins and losses.
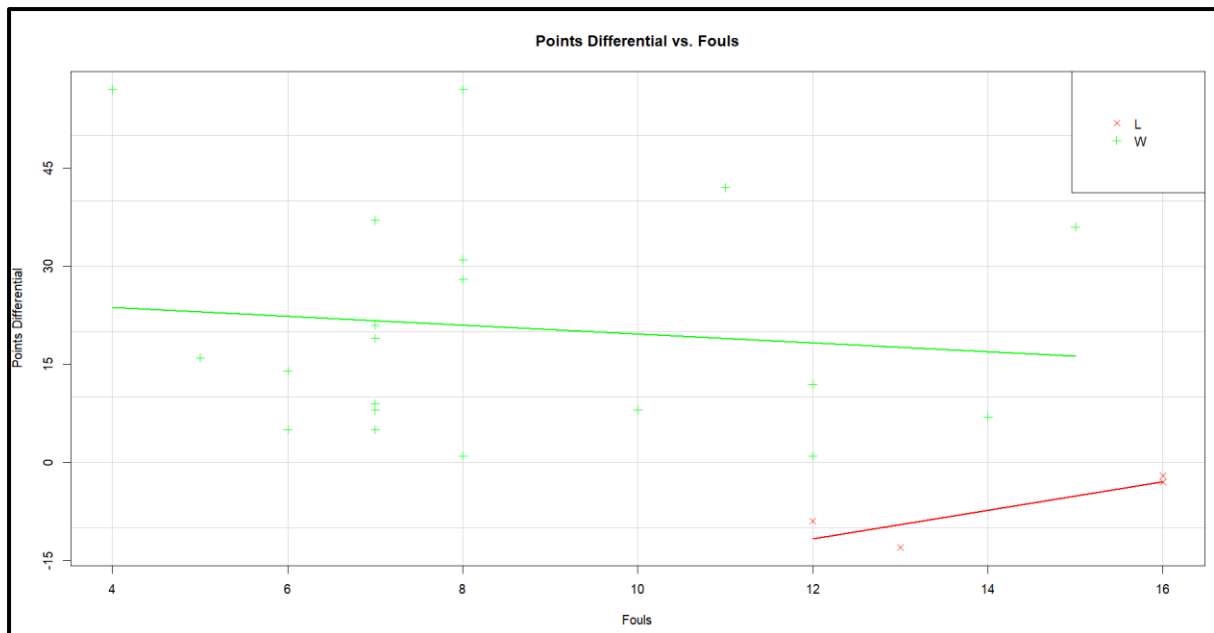
Graph 7: Points differential vs. number of fouls, grouped by result.

Evidently, there is an obvious relationship between the number of fouls committed by Slay-Ups and the result of the game. From graph 6, the average fouls committed in a loss (14.25) is far greater than in a win (8.45). Additionally, graph 7 demonstrates a decreasing relationship between points differential and fouls, with increasing number of fouls results in a smaller win margin. Although at first glance this relationship is the opposite for losses. However, the data points at points differential of -2 and 16 fouls was a game that went to overtime, hence increased number of minutes, leading to greater number of fouls. Additionally, the sample size for losses is only 4 games, hence it is impractical to try to draw any conclusion from this relationship.

From this analysis, it could be deduced that the number of fouls is an influential factor in the success of Slay-Ups.

3.1.5   Predictive Modelling

Using the team data within the "BBALL_STATS.csv" file, a variety of predictive models were

hypothesis tested to try and predict which variables were most influential on the success of Slay-Ups.

The models hypothesised and the statistically significant variables were as follows (table 4).

Model	Variables	Statistically Significant Variables
Two-Way	- qtr1_for - qtr2_for - qtr3_for - qtr4_for - qtr1_against - qtr2_against - qtr3_against - qtr4_against - fouls	- qtr2_against (5%) - fouls (5%) - Intercept (1%)
Offensive	- qtr1_for - qtr2_for - qtr3_for - qtr4_for - fouls	- fouls (1%) - Intercept (0.1%)
Defensive	- qtr1_against - qtr2_against - qtr3_against - qtr4_against - fouls	- qtr2_against (1%) - qtr3_against (5%) - fouls (1%) - Intercept (0.1%)
First Half	- qtr1_for - qtr2_for - qtr1_against - qtr2_against	- qtr2_against (0.1%) - Intercept (1%)
Second Half	- qtr3_for - qtr4_for - qtr3_against - qtr4_against	- qtr3_against (10%) - Intercept (1%)
First Quarter	- qtr1_for - qtr1_against	Intercept (5%)
Second Quarter	- qtr2_for - qtr2_against	- qtr2_against (0.1%) - Intercept (0.1%)
Third Quarter	- qtr3_for - qtr3_against	- qtr3_against (10%) - Intercept (1%)
Fourth Quarter	- qtr4_for - qtr4_against	Intercept (1%)
Best	- qtr2_against - qtr3_against - fouls	- qtr2_against (1%) - qtr3_against (5%) - fouls (0.1%) - Intercept (0.1%)

Table 4: Model Hypothesis Testing

Using the previous models, the "Best" model was theorised to be the most accurate predictor of success based upon the known data.

Hence, it can be deduced that the most influential factors on winning were the points scored against in the 2nd and 3rd quarters and the number fouls committed.

3.2 By Minute Team Statistics

3.3 By Minute Player Statistics

3.4 By Minute Lineup Statistics

4.0 Discussion

5.0 Limitations

6.0 Conclusion