Can algorithms play Moneyball?

Zachary Agrue, Patricia Mills, and Charles Vanleuvan

June 10, 2020

# Introduction

The 2011 film, Moneyball, tells the story of how Billy Beane and Paul DePodesta found a way to help the Oakland A’s improve their team with little money and a lot of data. Up to the A’s in 2002, players’ salaries and trade value were determined almost exclusively on the player’s batting average, RBIs, and homeruns with a dose of “gut feeling” from the scouts and coaches. That year, the Yankees players earned a combined 125,928,583. The A’s, by comparison, earned 39,679,746. Both teams were eliminated in the first round of the playoffs. Enter Sabermetrics.

Paul DePodesta did not think in terms of buy players but to buy runs. The three players that the A’s got in the off season that year were sought out because they got on base, which is the first step to getting runs. This has completely changed the world of baseball over the 18 years since the A’s did it.

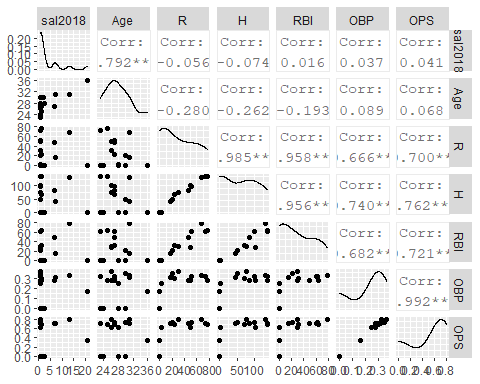
In 2018, the Chicago White Sox lost 100 games. In 2019, they lost 89. While still a losing season, the team made some changes in the off season. Meanwhile, in 2018 the Washington Nationals were barely above .500 but in 2019, they won the World Series. What did these teams change? Can machine learning algorithms identify who they should trade in order to play a little Moneyball?

# Analysis and Models

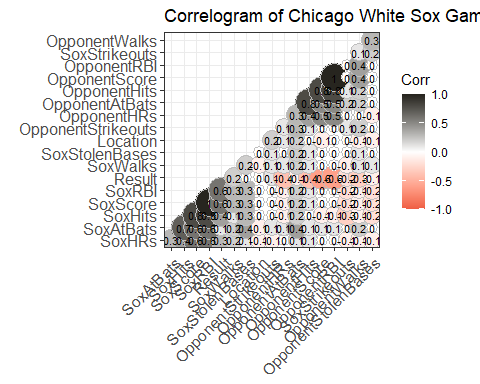
## About the Data

Scatterplots are key visual tools that describe relationships between two categories. In this report, pitching and hitting metrics are compared against player salary. The fundamental relationship that needs to be understood is player performance and player salary. Anecdotal evidence suggests that the more a player is paid, the better their performance is.

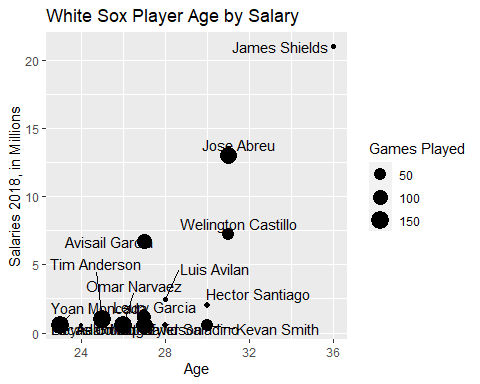
From the scatter plot matrix, pairs of features that show strong relationships can be observed.



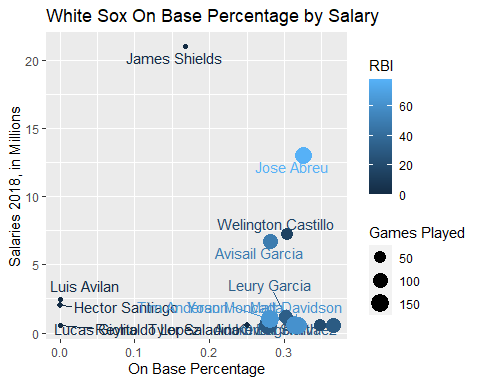
When considering game level data, the correlogram below shows high correlation for several statistics for the Chicago White Sox. Some are obvious correlations like the positive correlation between OpponentRBI and OpponentScore. Others seem less clear such as the slightly negative correlation between SoxHits and SoxStrikeouts. Interestingly, there is not the same correlation for OpponentHits and OpponentStrikeouts.



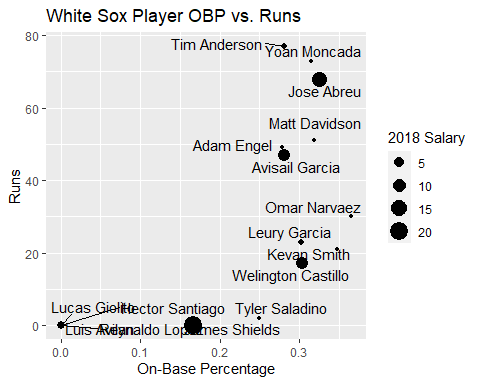
Older players are more established in the league and typically can negotiate higher salary on previous years’ performance. However, age shows a weak correlation (max = 0.262 R) to most offensive statistics. The average salary of players over 28 years old is greater than 5 million, yet the average salary for players under 28 years is below 2 million. Salary does not have a R score greater than 0.1 for any of the offensive stats.



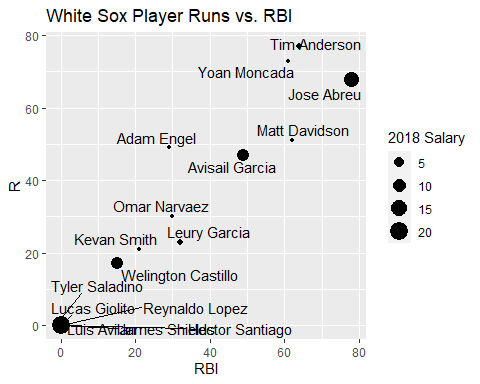
Excluding pitchers, the OBP for position players does not scale well with salary. Of the top 3 salary earners on the team, only 2 of those players reside in the top 5 players with highest OBP. On base percentage is a record of how often a player reaches base safely. In theory, these on base events should translate to more Runs. A player that scores many runs is directly contributing to a team’s score and increases the chance of winning.



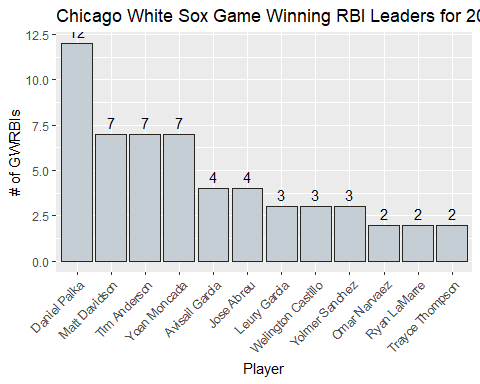
On the White Sox, it isn’t necessarily true that higher OBP scales with runs. The R value for OBP by runs is 0.67, and yet the Tim Anderson, one of the lowest paid players, has the highest runs scored while nearly having the smallest OBP of the position players. If OBP is not a strong predictor for runs scored, then RBI might be.



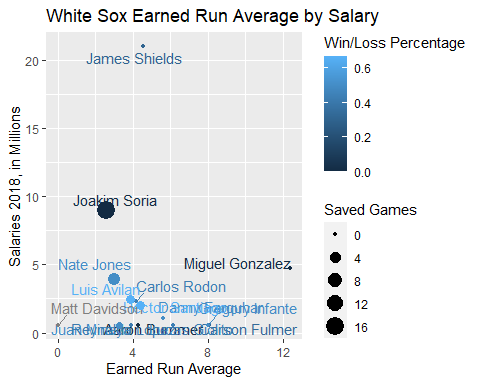
A strong R value (R = 0.958) between runs and RBI indicates that players who hit in more runs also individually contribute to the teams score by reaching homeplate safely themselves. Jose Abreu is the teams highest paid position player, has the most RBIs, and is third in runs scored. Two other players have similar performance to him, Tim Anderson and Yoan Moncada, yet they are paid significantly less than Abreu.



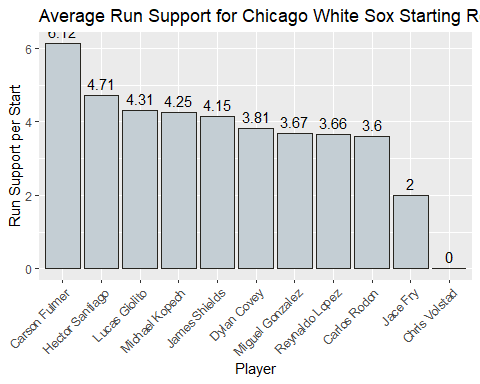
Game winning RBIs are runs that a player hits in that end up being the winning run for the game. In this study, this can be looked at as the “clutch performance factor”. Certain players show the tendency to perform well given a game-deciding situation. Daniel Palka, with 12 runs batted in, has approximately 40% more GWRBI’s than his closest teammate. This tendency can be a strong influencing factor in deciding to keep a player or update his contract.



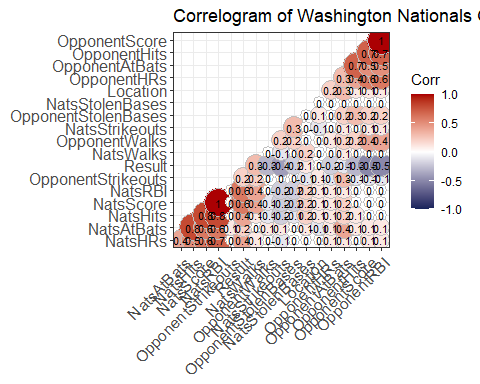
A pitcher’s ERA is the gold standard for comparison between pitchers, where lower ERA is better. It is the total number of earned runs averaged out for every 9 innings pitched. The correlation between ERA and Salary suggests that ERA is a strong influence in a pitcher’s salary. Barring outliers Matt Davidson and Miguel Gonzalez (G = 3), as ERA decreases, the pitcher salary increases. The same relationship is exhibited by salary and the number of saved games.



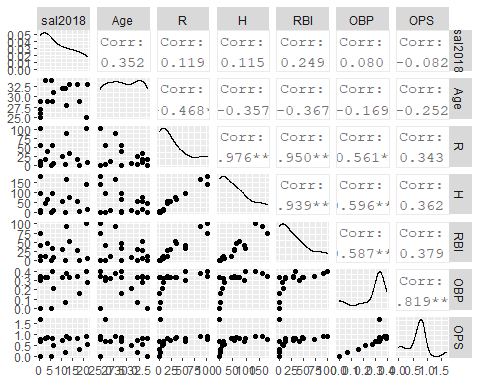
The Chicago White Sox pitching staff depend upon the batters to give run support to cover the pitcher’s earned runs. While Carson Fulmer’s run support average is over 6 runs per start, it does not counteract his Earned Run Average of 8.07. In fact, for the Sox pitching staff, Hector Santiago is the only pitcher whose average run support covers his ERA of 4.4, which explains why he has the highest win-loss percentage of the starting pitchers at 67%.



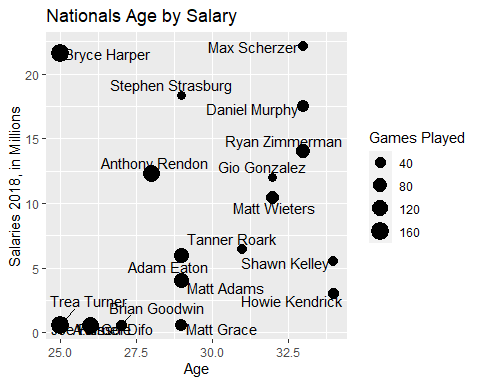
The correlogram of the Washington Nationals’ game level statistics shows that there is a slight positive correlation between NatsWalks and NatsScore. With the rest of the correlations, there is not much that is not to be expected in baseball.



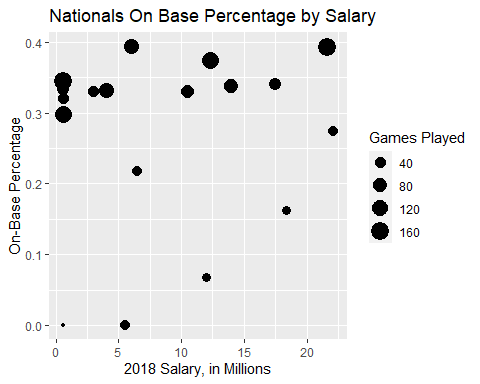
Similar to the White Sox, there is only a weak correlation between salary and most major offensive categories (R < 0.25). Stronger correlations exist within the offensive categories themselves (R > 0.95)



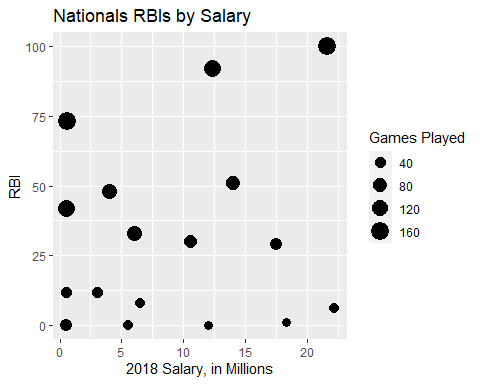
Another consistency between the two ball clubs is that older players have higher salaries, without necessarily providing better offense. Trea Turner’s offensive stats were similar to Bryce Harper but was paid nearly 20 times less than Bryce.



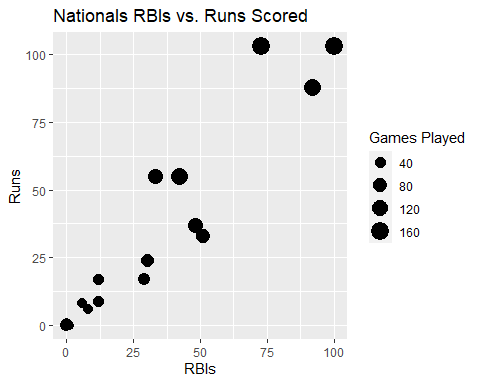
The Nationals have a higher OBP across their player salary range compared to the White Sox because they were a better offensive team. Still, however, the correlation between salary and OBP is R < 0.1.



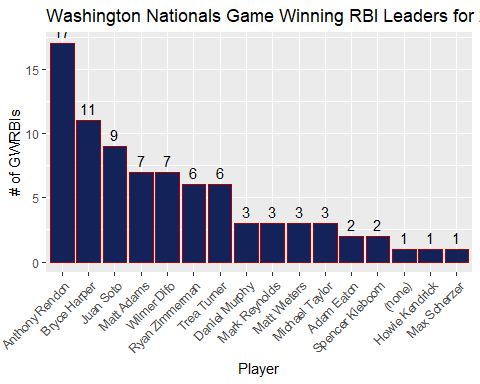
Runs batted in for the Nationals team show that the top three leaders in RBIs also played in the most games. However, it does not correlate to salary for all three. Two of the three are in the are amongst the high salary group while one is at or near the league minimum salary for 2018.



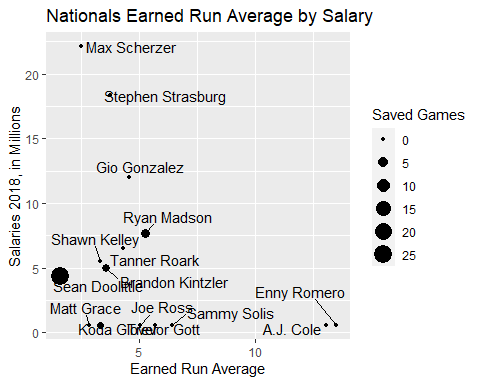
Like the White Sox, the Nationals players have a strong correlation between batting in runs and scoring runs (R = 0.95). Again, runs scored directly contribute to the teams score and thus increase the odds of winning the game. A player’s RBIs should be taken into consideration by General Manager’s deciding the team roster.



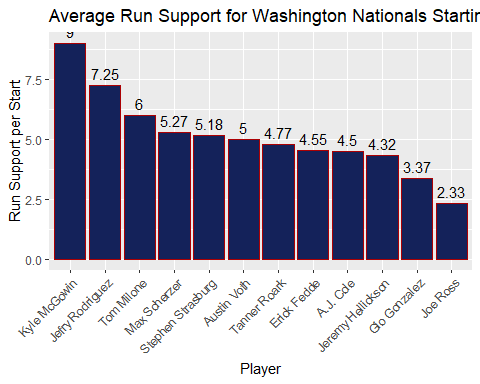
In terms of game winning performances, Anthony Rendon (3B) is paid almost $10 million less than Daniel Murphy (2B) but has 2.5 times the amount of game winning RBIs than Murphy (when averaged out for Game Winning RBIs per Game). Rendon also is 100 points higher than Murphy for OPS (On-base percentage plus slugging).



The Nationals starting rotation exhibit a clear and direct relationship between ERA and Salary for starting pitchers. Nationals pitchers with lower ERA in 2018 also had higher salaries. As with the White Sox pitchers, the ERA remains the gold standard in determining a pitcher’s salary.



Unlike the Chicago White Sox, the Washington Nationals pitching staff enjoys high run support for most of the starting pitchers. For Washington, Austin Voth (6.57 ERA), Erick Fedde (5.54 ERA), AJ Cole (13.06 ERA), Gio Gonzalez (4.57 ERA), and Joe Ross (5.06 ERA) are the pitchers who draw less run support than their ERAs. The majority of the starting pitchers have far higher run support than needed.



## Model 1

Association Rules Mining helped immensely to find trends for winning for each team. Two sets of rules were created for each team: winning games all data and winning batting lineups. For the all data rules, the parameters were set at a support of 15%, confidence of 50%, and a minimum length of 3. For the White Sox, this resulted in 42 rules. Meanwhile, 61 rules were generated for the Nationals.

The rules generated based on the batting lineups had parameters of support 15%, confidence 90%, and a minimum length of 3. This generated 24 rules for the White Sox and 103 rules for the Nationals.

## Model 2

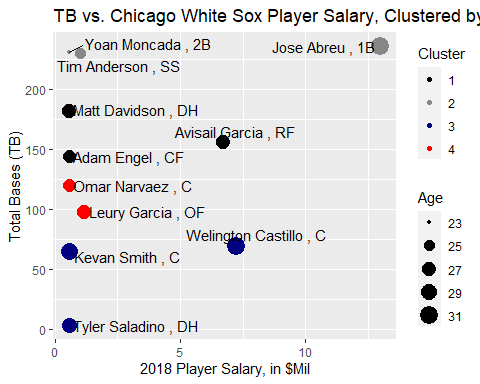
In order to determine which players were either being overpaid or underpaid by their team, players’ salaries were grouped as low, mid, or high based on the mean salary for batters or pitchers for their team, based on position played. Then using random forest, each player’s salary group was predicted based on comparisons to their batting or pitching stats, as appropriate.

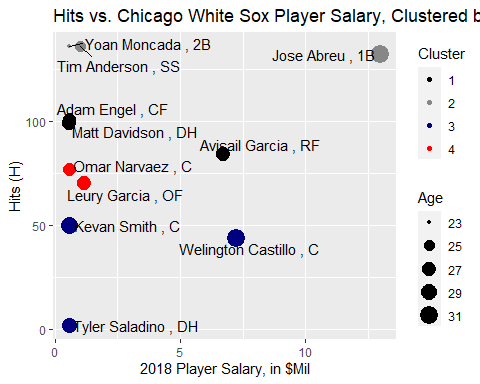
Holdout testing was used to create as accurate a model as possible. Four folds were used for both team’s batters. Five folds were used for pitchers. These folds resulted in groups of 4 players per holdout. Then a loop was written to create new test and train data frames. Each training group used all of the data in the salary and stats data frames. Once classification was made, each player’s name, original salary range, and new salary range was added to a data frame of results for that team and that position – batters versus pitchers.

## Model 3

In order to optimize k value, the elbow method, looped through k values 1 to 10 which returns a rather interesting plot. The elbow where the Sum of Squared errors can be interpreted as any k value between k = 2 or k = 6, based on the shape of the plot.

## Rk Pos Player Age G PA AB R H X2B X3B HR RBI SB CS BB SO  
## 1 1 C Omar Narvaez 26 97 322 280 30 77 14 1 9 30 0 2 38 65  
## 2 2 1B Jose Abreu 31 128 553 499 68 132 36 1 22 78 2 0 37 109  
## 3 3 2B Yoan Moncada 23 149 650 578 73 136 32 6 17 61 12 6 67 217  
## 4 4 SS Tim Anderson 25 153 606 567 77 136 28 3 20 64 26 8 30 149  
## 5 5 3B Yolmer Sanchez 26 155 662 600 62 145 34 10 8 55 14 6 49 138  
## 6 6 LF Nicky Delmonico 25 88 318 284 31 61 11 5 8 25 1 2 27 80  
## 7 7 CF Adam Engel 26 143 463 429 49 101 17 4 6 29 16 8 18 129  
## 8 8 RF Avisail Garcia 27 93 385 356 47 84 11 2 19 49 3 1 20 102  
## 9 9 DH Matt Davidson 27 123 496 434 51 99 23 0 20 62 0 0 52 165  
## 10 10 OF Daniel Palka 26 124 449 417 56 100 15 3 27 67 2 1 30 153  
## 11 11 OF Leury Garcia 27 82 275 258 23 70 7 4 4 32 12 1 9 69  
## 12 12 C Kevan Smith 30 52 187 171 21 50 6 0 3 21 1 0 10 18  
## 13 13 C Welington Castillo 31 49 181 170 17 44 7 0 6 15 1 0 9 46  
## 14 14 RF Trayce Thompson 27 48 130 121 14 14 3 0 3 9 3 1 7 46  
## 15 15 LF Charlie Tilson 25 41 121 106 7 28 1 1 0 11 2 3 10 20  
## 16 16 IF Jose Rondon 24 42 107 100 15 23 6 0 6 14 2 1 7 30  
## 17 17 LF Ryan LaMarre 29 33 71 66 8 20 6 0 2 10 1 1 2 20  
## 18 18 OF Ryan Cordell 26 19 40 37 3 4 1 0 1 4 0 0 0 15  
## 19 19 DH Matt Skole 28 4 13 11 2 3 0 0 1 1 0 0 2 3  
## 20 20 C Alfredo Gonzalez 25 3 9 9 0 1 0 0 0 1 0 0 0 4  
## 21 21 DH Tyler Saladino 28 6 9 8 2 2 1 0 0 0 0 0 0 3  
## 22 22 C Dustin Garneau 30 1 3 2 0 1 0 0 0 1 0 0 1 0  
## 23 23 P James Shields 36 3 7 6 0 1 0 0 0 0 0 0 0 3  
## 24 24 P Lucas Giolito 23 3 6 6 0 0 0 0 0 0 0 0 0 4  
## 25 25 P Hector Santiago 30 5 4 4 0 0 0 0 0 0 0 0 0 2  
## 26 26 P Dylan Covey 26 1 1 1 0 0 0 0 0 0 0 0 0 1  
## 27 27 P Chris Beck 27 4 1 1 0 0 0 0 0 0 0 0 0 1  
## 28 28 P Luis Avilan 28 3 1 1 0 0 0 0 0 0 0 0 0 1  
## 29 29 P Reynaldo Lopez 24 1 1 1 0 0 0 0 0 0 0 0 0 1  
## BA OBP SLG OPS OPS. TB GDP HBP SH SF IBB BATS  
## 1 0.275 0.366 0.429 0.794 119 120 5 2 2 0 1 LEFT  
## 2 0.265 0.325 0.473 0.798 117 236 14 11 0 6 7 RIGHT  
## 3 0.235 0.315 0.400 0.714 96 231 4 1 2 2 1 SWITCH  
## 4 0.240 0.281 0.406 0.687 87 230 15 4 2 3 2 RIGHT  
## 5 0.242 0.306 0.372 0.678 87 223 9 8 2 3 0 SWITCH  
## 6 0.215 0.296 0.373 0.669 84 106 7 6 0 1 1 LEFT  
## 7 0.235 0.279 0.336 0.614 69 144 1 8 7 1 0 RIGHT  
## 8 0.236 0.281 0.438 0.719 95 156 9 4 0 5 2 RIGHT  
## 9 0.228 0.319 0.419 0.738 102 182 8 7 0 3 0 RIGHT  
## 10 0.240 0.294 0.484 0.778 110 202 7 2 0 0 3 LEFT  
## 11 0.271 0.303 0.376 0.679 87 97 2 3 4 1 0 SWITCH  
## 12 0.292 0.348 0.380 0.728 102 65 5 5 0 1 0 RIGHT  
## 13 0.259 0.304 0.406 0.710 94 69 7 2 0 0 0 RIGHT  
## 14 0.116 0.163 0.215 0.378 4 26 0 0 1 1 1 RIGHT  
## 15 0.264 0.331 0.292 0.623 75 31 2 1 2 1 0 LEFT  
## 16 0.230 0.280 0.470 0.750 102 47 3 0 0 0 0 RIGHT  
## 17 0.303 0.324 0.485 0.809 120 32 0 1 0 2 0 RIGHT  
## 18 0.108 0.125 0.216 0.341 -8 8 0 1 0 2 0 RIGHT  
## 19 0.273 0.385 0.545 0.930 154 6 0 0 0 0 0 LEFT  
## 20 0.111 0.111 0.111 0.222 -38 1 0 0 0 0 0 RIGHT  
## 21 0.250 0.250 0.375 0.625 70 3 0 0 1 0 0 RIGHT  
## 22 0.500 0.667 0.500 1.167 232 1 0 0 0 0 0 RIGHT  
## 23 0.167 0.167 0.167 0.333 -7 1 0 0 1 0 0 RIGHT  
## 24 0.000 0.000 0.000 0.000 -100 0 0 0 0 0 0 RIGHT  
## 25 0.000 0.000 0.000 0.000 -100 0 1 0 0 0 0 RIGHT  
## 26 0.000 0.000 0.000 0.000 -100 0 0 0 0 0 0 RIGHT  
## 27 0.000 0.000 0.000 0.000 -100 0 0 0 0 0 0 RIGHT  
## 28 0.000 0.000 0.000 0.000 -100 0 0 0 0 0 0 LEFT  
## 29 0.000 0.000 0.000 0.000 -100 0 0 0 0 0 0 RIGHT

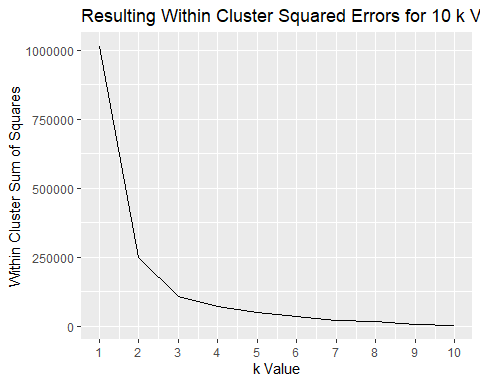


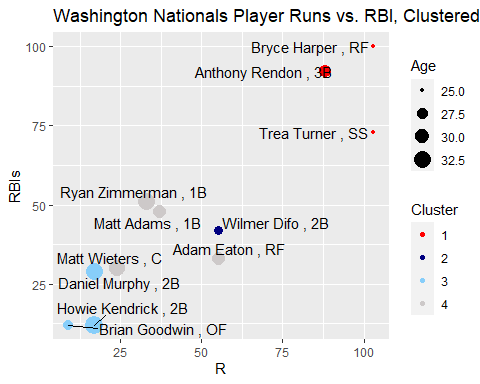


Using K-means clustering, we can group the players with similar performance together, and use the clusters to compare players. Using the elbow method, we can observe where the Within-Cluster Sum of Squares first starts to diminish, forming an “elbow” in the plot of k vs. WCSoS. From this, we see that the optimal k value is 3. When given business context, this can be construed to make sense. There are three categories of positions players: Outfielders, Infielders, and Catchers.

#clustering players  
WashingtonNationals2018Batting # WAS batting stats for 2018

## Rk Pos Player Age G PA AB R H X2B X3B HR RBI SB CS BB  
## 1 1 C Matt Wieters 32 76 271 235 24 56 8 0 8 30 0 1 30  
## 2 2 1B Ryan Zimmerman 33 85 323 288 33 76 21 2 13 51 1 1 30  
## 3 3 2B Wilmer Difo 26 148 456 408 55 94 14 7 7 42 10 3 39  
## 4 4 SS Trea Turner 25 162 740 664 103 180 27 6 19 73 43 9 69  
## 5 5 3B Anthony Rendon 28 136 597 529 88 163 44 2 24 92 2 1 55  
## 6 6 LF Juan Soto 19 116 494 414 77 121 25 1 22 70 5 2 79  
## 7 7 CF Michael A. Taylor 27 134 385 353 46 80 22 3 6 28 24 6 29  
## 8 8 RF Bryce Harper 25 159 695 550 103 137 34 0 34 100 13 3 130  
## 9 9 RF Adam Eaton 29 95 370 319 55 96 18 1 5 33 9 1 38  
## 10 10 1B Matt Adams 29 94 277 249 37 64 9 0 18 48 0 0 24  
## 11 11 1B Mark Reynolds 34 86 235 206 26 51 8 0 13 40 0 0 24  
## 12 12 C Pedro Severino 24 70 213 190 14 32 9 0 2 15 1 0 18  
## 13 13 2B Daniel Murphy 33 56 205 190 17 57 9 0 6 29 1 0 13  
## 14 14 2B Howie Kendrick 34 40 160 152 17 46 14 0 4 12 1 1 5  
## 15 15 C Spencer Kieboom 27 52 143 125 16 29 5 0 2 13 0 0 16  
## 16 16 OF Andrew Stevenson 24 57 86 75 9 19 2 0 1 13 1 1 6  
## 17 17 OF Brian Goodwin 27 48 79 65 9 13 1 0 3 12 3 1 10  
## 18 18 OF Victor Robles 21 21 66 59 8 17 3 1 3 10 3 2 4  
## 19 19 OF Moises Sierra 29 27 60 54 4 9 2 0 0 4 1 1 2  
## 20 20 IF Adrian Sanchez 27 28 59 58 8 16 2 1 0 3 0 0 1  
## 21 21 3B Matt Reynolds 27 12 14 13 1 2 0 0 0 1 0 0 1  
## 22 22 C Miguel Montero 34 4 13 11 0 0 0 0 0 0 0 0 2  
## 23 23 OF Rafael Bautista 25 9 6 6 1 0 0 0 0 0 0 0 0  
## 24 24 P Max Scherzer 33 32 78 70 8 17 2 0 0 6 1 0 1  
## 25 25 P Tanner Roark 31 29 65 58 6 11 2 1 0 8 0 0 1  
## 26 26 P Stephen Strasburg 29 22 51 41 0 5 0 0 0 1 0 0 2  
## 27 27 P Gio Gonzalez 32 24 47 44 1 3 1 0 0 0 0 0 0  
## 28 28 P Jeremy Hellickson 31 18 35 32 0 2 1 0 0 1 0 0 0  
## 29 29 P Jefry Rodriguez 24 14 18 16 2 3 1 0 0 1 0 0 0  
## 30 30 P Erick Fedde 25 10 17 16 1 1 0 0 0 0 0 0 1  
## 31 31 P Tommy Milone 31 5 9 7 0 0 0 0 0 0 0 0 1  
## 32 32 P Joe Ross 25 3 5 5 0 0 0 0 0 0 0 0 0  
## 33 33 P A.J. Cole 26 4 4 3 1 1 0 0 1 1 0 0 0  
## 34 34 P Wander Suero 26 38 3 3 0 0 0 0 0 0 0 0 0  
## 35 35 P Matt Grace 29 54 3 3 0 1 0 0 0 0 0 0 0  
## 36 36 P Kyle McGowin 26 5 2 2 1 0 0 0 0 0 0 0 0  
## 37 37 P Austin Voth 26 4 2 2 0 0 0 0 0 0 0 0 0  
## 38 38 P Shawn Kelley 34 32 1 1 0 0 0 0 0 0 0 0 0  
## 39 39 P Justin Miller 31 46 1 1 0 0 0 0 0 0 0 0 0  
## SO BA OBP SLG OPS OPS. TB GDP HBP SH SF IBB BATS  
## 1 45 0.238 0.330 0.374 0.704 86 88 5 3 1 2 3 SWITCH  
## 2 55 0.264 0.337 0.486 0.824 114 140 10 3 0 2 1 RIGHT  
## 3 82 0.230 0.298 0.350 0.649 71 143 8 2 3 4 5 SWITCH  
## 4 132 0.271 0.344 0.416 0.760 100 276 7 5 2 0 3 RIGHT  
## 5 82 0.308 0.374 0.535 0.909 137 283 5 5 0 8 5 RIGHT  
## 6 99 0.292 0.406 0.517 0.923 142 214 9 0 1 0 10 LEFT  
## 7 116 0.227 0.287 0.357 0.644 69 126 9 1 2 0 2 RIGHT  
## 8 169 0.249 0.393 0.496 0.889 133 273 7 6 0 9 16 LEFT  
## 9 64 0.301 0.394 0.411 0.805 114 131 2 11 2 0 0 LEFT  
## 10 55 0.257 0.332 0.510 0.842 118 127 6 4 0 0 2 LEFT  
## 11 64 0.248 0.328 0.476 0.803 109 98 8 2 0 3 1 RIGHT  
## 12 47 0.168 0.254 0.247 0.501 34 47 3 4 0 1 4 RIGHT  
## 13 17 0.300 0.341 0.442 0.784 105 84 4 0 0 2 2 LEFT  
## 14 29 0.303 0.331 0.474 0.805 110 72 6 2 0 1 1 RIGHT  
## 15 28 0.232 0.322 0.320 0.642 71 40 2 1 0 1 0 RIGHT  
## 16 23 0.253 0.306 0.320 0.626 66 24 0 1 1 3 0 LEFT  
## 17 26 0.200 0.321 0.354 0.674 78 23 0 2 1 1 0 LEFT  
## 18 12 0.288 0.348 0.525 0.874 127 31 2 2 0 1 0 RIGHT  
## 19 20 0.167 0.217 0.204 0.420 12 11 2 2 0 2 0 RIGHT  
## 20 8 0.276 0.288 0.345 0.633 67 20 0 0 0 0 0 RIGHT  
## 21 4 0.154 0.214 0.154 0.368 0 2 0 0 0 0 0 RIGHT  
## 22 3 0.000 0.154 0.000 0.154 -54 0 0 0 0 0 1 LEFT  
## 23 1 0.000 0.000 0.000 0.000 -100 0 1 0 0 0 0 RIGHT  
## 24 14 0.243 0.274 0.271 0.545 45 19 1 2 5 0 0 RIGHT  
## 25 19 0.190 0.217 0.259 0.475 25 15 1 1 5 0 0 RIGHT  
## 26 12 0.122 0.163 0.122 0.285 -23 5 3 0 8 0 0 RIGHT  
## 27 27 0.068 0.068 0.091 0.159 -58 4 0 0 3 0 0 RIGHT  
## 28 13 0.063 0.063 0.094 0.156 -59 3 0 0 3 0 0 RIGHT  
## 29 8 0.188 0.188 0.250 0.438 14 4 0 0 2 0 0 RIGHT  
## 30 5 0.063 0.118 0.063 0.180 -50 1 1 0 0 0 0 RIGHT  
## 31 3 0.000 0.125 0.000 0.125 -63 0 0 0 1 0 0 LEFT  
## 32 3 0.000 0.000 0.000 0.000 -100 0 1 0 0 0 0 RIGHT  
## 33 1 0.333 0.333 1.333 1.667 311 4 0 0 1 0 0 RIGHT  
## 34 1 0.000 0.000 0.000 0.000 -100 0 0 0 0 0 0 RIGHT  
## 35 0 0.333 0.333 0.333 0.667 77 1 0 0 0 0 0 LEFT  
## 36 0 0.000 0.000 0.000 0.000 -100 0 0 0 0 0 0 RIGHT  
## 37 1 0.000 0.000 0.000 0.000 -100 0 0 0 0 0 0 RIGHT  
## 38 1 0.000 0.000 0.000 0.000 -100 0 0 0 0 0 0 RIGHT  
## 39 0 0.000 0.000 0.000 0.000 -100 0 1 0 0 0 0 RIGHT





## Model 4

kMeans clustering provided a method for unsupervised pattern recognition within the White Sox and Nationals datasets. There are also supervised learning methods that utilize the same basic principles of determining similarity between data points based on distance between all vectors of the examples. One such method is kNN, which is an abbreviation for k-Nearest Neighbors. The term “k” acts in kNN similar to the “k” in kMeans, where it can be arbitrarily decided by the model user. In kNN, k represents the number of neighbors, or closest data points, to a test record example that will be used to calculate the majority classification voting that will determine the predicted class. If k is small, then the model is very sensitive to noise, and if k is large, then the neighborhood is too large and the estimated class will be generalized. A unique feature of kNN modeling is there is no standard training and testing split. The algorithm function takes both a training and testing input simultaneously, and the user denotes the training data set’s classification labels.

To run kNN modeling, all players needed to be classified based on their salary in three distinct groups. Players with salaries under 3 million in 2018 were labeled as “low”, players with salaries over $11 million in 2018 were labelled as “high”, and the rest were labeled as “mid”. The goal in this analysis is to find the number k that provides the best accuracy in the kNN model, when attempting to label players salaries based on their performance.

The Nationals and White Sox player data sets were relatively small. To create more data, sampling with replacement was done to fabricate more data that was identical to the original data set. This method of sampling simulated a player being compared to different teammates in different sampling cycles, so it strained the model to provide consistent accuracy feedback.

The process for running kNN with the player datasets was an iterative one. After identifying the optimal k value for batters and pitchers data sets, 3-fold cross validation was repeated and the accuracy score was calculated for each iteration.

## [1] 16

## Warning in split.default(sample(1:nrow(CWSsalbatstats18reduced2)), 1:k\_folds):  
## data length is not a multiple of split variable

## [1] "K-Fold Iteration: 1 Accuracy: 0.491"  
## [1] "K-Fold Iteration: 2 Accuracy: 0.212"  
## [1] "K-Fold Iteration: 3 Accuracy: 0.35"

## [1] 15

## [1] "K-Fold Iteration: 1 Accuracy: 0.208"  
## [1] "K-Fold Iteration: 2 Accuracy: 0.215"  
## [1] "K-Fold Iteration: 3 Accuracy: 0.589"

The Washington Nationals batter’s dataset was used to determine the optimal k value for batters. Using 3-fold cross validation to collect accuracy scores, k values were iterated from 1 to 5, and the accuracies plotted below. K was set to 5 for model testing as the accuracy plateaued for values larger than 5.

## [1] 19

## Warning in split.default(sample(1:nrow(WASsalbatstats18reduced2)), 1:k\_folds):  
## data length is not a multiple of split variable

## [1] "K-Fold Iteration: 1 Accuracy: 0.462"  
## [1] "K-Fold Iteration: 2 Accuracy: 0.489"  
## [1] "K-Fold Iteration: 3 Accuracy: 0.142"

## [1] 15

## [1] "K-Fold Iteration: 1 Accuracy: 0.598"  
## [1] "K-Fold Iteration: 2 Accuracy: 0.199"  
## [1] "K-Fold Iteration: 3 Accuracy: 0.8"

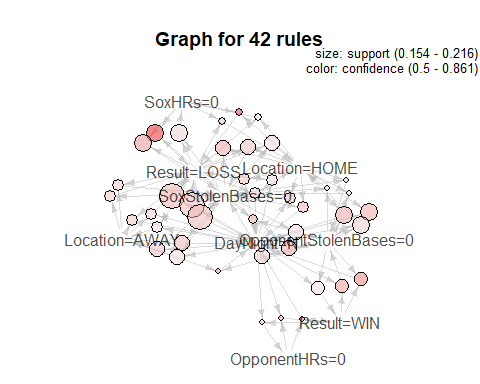
# Results

## Model 1

A couple of the more interesting rules generated for the White Sox is if they’re playing at night, and the opponents don’t hit any home runs (or have any stolen bases), they are more likely to win, as well as if they’re playing a home game and don’t hit any home runs themselves, they’re more likely to lose. This gives insight as to how schedule factors into overall performance, as well as how important defense is. But those rules don’t give any insight to player performance, so we generated rules based on the winning batting lineup of each game.

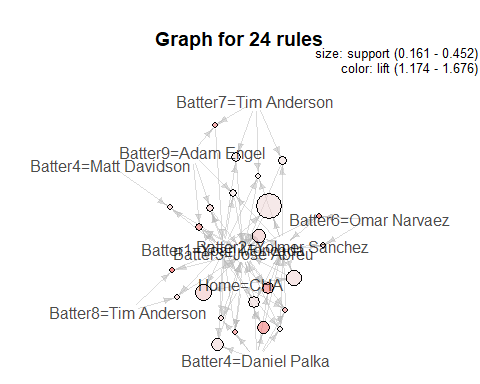
## lhs rhs support   
## [1] {SoxHRs=0,SoxStolenBases=0} => {Result=LOSS} 0.1913580  
## [2] {Location=HOME,SoxHRs=0} => {Result=LOSS} 0.1604938  
## [3] {Result=WIN,OpponentHRs=0} => {DayNight=N} 0.1543210  
## [4] {Result=WIN,OpponentStolenBases=0} => {DayNight=N} 0.1790123  
## [5] {Result=LOSS,SoxHRs=0} => {SoxStolenBases=0} 0.1913580  
## [6] {DayNight=N,OpponentHRs=0} => {Result=WIN} 0.1543210  
## confidence lift count  
## [1] 0.8611111 1.395000 31   
## [2] 0.8387097 1.358710 26   
## [3] 0.7352941 1.215486 25   
## [4] 0.7250000 1.198469 29   
## [5] 0.7045455 1.214217 31   
## [6] 0.6944444 1.814516 25

## lhs rhs support confidence lift count  
## [1] {SoxHRs=0,   
## SoxStolenBases=0} => {Result=LOSS} 0.1913580 0.8611111 1.3950000 31  
## [2] {Location=HOME,   
## SoxHRs=0} => {Result=LOSS} 0.1604938 0.8387097 1.3587097 26  
## [3] {Result=WIN,   
## OpponentHRs=0} => {DayNight=N} 0.1543210 0.7352941 1.2154862 25  
## [4] {Result=WIN,   
## OpponentStolenBases=0} => {DayNight=N} 0.1790123 0.7250000 1.1984694 29  
## [5] {Result=LOSS,   
## SoxHRs=0} => {SoxStolenBases=0} 0.1913580 0.7045455 1.2142166 31  
## [6] {DayNight=N,   
## OpponentHRs=0} => {Result=WIN} 0.1543210 0.6944444 1.8145161 25  
## [7] {Location=AWAY,   
## OpponentStolenBases=0} => {DayNight=N} 0.1543210 0.6944444 1.1479592 25  
## [8] {Result=WIN,   
## DayNight=N} => {OpponentStolenBases=0} 0.1790123 0.6904762 1.3476764 29  
## [9] {Location=HOME,   
## SoxStolenBases=0} => {Result=LOSS} 0.1851852 0.6666667 1.0800000 30  
## [10] {Location=HOME,   
## OpponentStolenBases=0} => {DayNight=N} 0.1913580 0.6595745 1.0903170 31  
## [11] {Location=HOME,   
## DayNight=N} => {OpponentStolenBases=0} 0.1913580 0.6595745 1.2873622 31  
## [12] {SoxStolenBases=0,   
## OpponentStolenBases=0} => {DayNight=N} 0.1851852 0.6382979 1.0551455 30  
## [13] {Result=LOSS,   
## OpponentStolenBases=0} => {DayNight=N} 0.1666667 0.6279070 1.0379687 27  
## [14] {Result=LOSS,   
## DayNight=N} => {SoxStolenBases=0} 0.2160494 0.6250000 1.0771277 35  
## [15] {Location=HOME,   
## SoxStolenBases=0} => {DayNight=N} 0.1728395 0.6222222 1.0285714 28  
## [16] {Location=AWAY,   
## SoxStolenBases=0} => {DayNight=N} 0.1851852 0.6122449 1.0120783 30  
## [17] {DayNight=N,   
## SoxStolenBases=0} => {Result=LOSS} 0.2160494 0.6034483 0.9775862 35  
## [18] {Result=LOSS,   
## SoxStolenBases=0} => {DayNight=N} 0.2160494 0.6034483 0.9975369 35  
## [19] {Location=HOME,   
## DayNight=N} => {SoxStolenBases=0} 0.1728395 0.5957447 1.0267089 28  
## [20] {Location=HOME,   
## DayNight=N} => {Result=LOSS} 0.1728395 0.5957447 0.9651064 28  
## [21] {Result=WIN,   
## DayNight=N} => {OpponentHRs=0} 0.1543210 0.5952381 1.8543956 25  
## [22] {Result=LOSS,   
## SoxHRs=0} => {Location=HOME} 0.1604938 0.5909091 1.1818182 26  
## [23] {Location=AWAY,   
## DayNight=N} => {SoxStolenBases=0} 0.1851852 0.5882353 1.0137672 30  
## [24] {Location=HOME,   
## Result=LOSS} => {SoxStolenBases=0} 0.1851852 0.5882353 1.0137672 30  
## [25] {Location=AWAY,   
## SoxStolenBases=0} => {Result=LOSS} 0.1728395 0.5714286 0.9257143 28  
## [26] {Location=AWAY,   
## Result=LOSS} => {SoxStolenBases=0} 0.1728395 0.5714286 0.9848024 28  
## [27] {Location=AWAY,   
## Result=LOSS} => {DayNight=N} 0.1728395 0.5714286 0.9446064 28  
## [28] {Location=HOME,   
## SoxStolenBases=0} => {OpponentStolenBases=0} 0.1543210 0.5555556 1.0843373 25  
## [29] {DayNight=N,   
## OpponentStolenBases=0} => {Location=HOME} 0.1913580 0.5535714 1.1071429 31  
## [30] {Location=AWAY,   
## DayNight=N} => {Result=LOSS} 0.1728395 0.5490196 0.8894118 28  
## [31] {Location=HOME,   
## Result=LOSS} => {DayNight=N} 0.1728395 0.5490196 0.9075630 28  
## [32] {DayNight=N,   
## OpponentStolenBases=0} => {SoxStolenBases=0} 0.1851852 0.5357143 0.9232523 30  
## [33] {Result=LOSS,   
## SoxStolenBases=0} => {SoxHRs=0} 0.1913580 0.5344828 1.5742947 31  
## [34] {Location=HOME,   
## OpponentStolenBases=0} => {SoxStolenBases=0} 0.1543210 0.5319149 0.9167044 25  
## [35] {SoxStolenBases=0,   
## OpponentStolenBases=0} => {Location=HOME} 0.1543210 0.5319149 1.0638298 25  
## [36] {DayNight=N,   
## OpponentStolenBases=0} => {Result=WIN} 0.1790123 0.5178571 1.3531106 29  
## [37] {DayNight=N,   
## SoxStolenBases=0} => {Location=AWAY} 0.1851852 0.5172414 1.0344828 30  
## [38] {Result=LOSS,   
## SoxStolenBases=0} => {Location=HOME} 0.1851852 0.5172414 1.0344828 30  
## [39] {DayNight=N,   
## SoxStolenBases=0} => {OpponentStolenBases=0} 0.1851852 0.5172414 1.0095555 30  
## [40] {Location=HOME,   
## Result=LOSS} => {SoxHRs=0} 0.1604938 0.5098039 1.5016043 26  
## [41] {Result=LOSS,   
## DayNight=N} => {Location=AWAY} 0.1728395 0.5000000 1.0000000 28  
## [42] {Result=LOSS,   
## DayNight=N} => {Location=HOME} 0.1728395 0.5000000 1.0000000 28



Those results were mostly centered around Jose Abreu being third in the lineup, but Adam Engel being ninth in the lineup, as well as Yoan Moncada being leadoff, and Yolmer Sanchez being second seem to be the only other consistencies withing the winning batting lineups. The more variable ones (the ones that didn’t have any rules generated about them), the players can be assessed to determine if their performance for the season merits them remaining on the team.

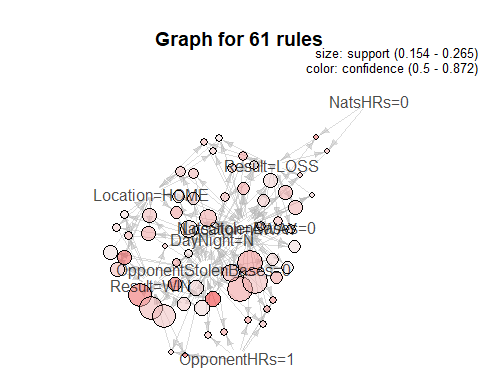
## lhs rhs support confidence lift count  
## [1] {Batter1=Yoan Moncada,   
## Batter8=Tim Anderson} => {Batter3=Jose Abreu} 0.1612903 1.0000000 1.291667 10  
## [2] {Batter3=Jose Abreu,   
## Batter8=Tim Anderson} => {Batter1=Yoan Moncada} 0.1612903 1.0000000 1.675676 10  
## [3] {Batter2=Yolmer Sanchez,   
## Batter6=Omar Narvaez} => {Batter3=Jose Abreu} 0.1612903 1.0000000 1.291667 10  
## [4] {Batter3=Jose Abreu,   
## Batter6=Omar Narvaez} => {Batter2=Yolmer Sanchez} 0.1612903 1.0000000 1.631579 10  
## [5] {Batter1=Yoan Moncada,   
## Batter4=Matt Davidson} => {Batter3=Jose Abreu} 0.1612903 1.0000000 1.291667 10  
## [6] {Home=CHA,   
## Batter1=Yoan Moncada} => {Batter3=Jose Abreu} 0.3225806 1.0000000 1.291667 20  
## [7] {Home=CHA,   
## Batter1=Yoan Moncada,   
## Batter4=Daniel Palka} => {Batter3=Jose Abreu} 0.1612903 1.0000000 1.291667 10  
## [8] {Home=CHA,   
## Batter1=Yoan Moncada,   
## Batter9=Adam Engel} => {Batter3=Jose Abreu} 0.1774194 1.0000000 1.291667 11  
## [9] {Home=CHA,   
## Batter1=Yoan Moncada,   
## Batter2=Yolmer Sanchez} => {Batter3=Jose Abreu} 0.2741935 1.0000000 1.291667 17  
## [10] {Home=CHA,   
## Batter2=Yolmer Sanchez} => {Batter3=Jose Abreu} 0.3064516 0.9500000 1.227083 19  
## [11] {Batter1=Yoan Moncada,   
## Batter4=Daniel Palka} => {Batter2=Yolmer Sanchez} 0.2580645 0.9411765 1.535604 16  
## [12] {Batter1=Yoan Moncada,   
## Batter4=Daniel Palka} => {Batter3=Jose Abreu} 0.2580645 0.9411765 1.215686 16  
## [13] {Batter1=Yoan Moncada,   
## Batter2=Yolmer Sanchez,   
## Batter4=Daniel Palka} => {Batter3=Jose Abreu} 0.2419355 0.9375000 1.210938 15  
## [14] {Batter1=Yoan Moncada,   
## Batter3=Jose Abreu,   
## Batter4=Daniel Palka} => {Batter2=Yolmer Sanchez} 0.2419355 0.9375000 1.529605 15  
## [15] {Batter1=Yoan Moncada,   
## Batter2=Yolmer Sanchez} => {Batter3=Jose Abreu} 0.4516129 0.9333333 1.205556 28  
## [16] {Batter1=Yoan Moncada,   
## Batter7=Tim Anderson} => {Batter3=Jose Abreu} 0.2096774 0.9285714 1.199405 13  
## [17] {Batter2=Yolmer Sanchez,   
## Batter7=Tim Anderson} => {Batter3=Jose Abreu} 0.1935484 0.9230769 1.192308 12  
## [18] {Home=CHA,   
## Batter3=Jose Abreu,   
## Batter9=Adam Engel} => {Batter1=Yoan Moncada} 0.1774194 0.9166667 1.536036 11  
## [19] {Batter1=Yoan Moncada,   
## Batter2=Yolmer Sanchez,   
## Batter9=Adam Engel} => {Batter3=Jose Abreu} 0.1774194 0.9166667 1.184028 11  
## [20] {Batter7=Tim Anderson,   
## Batter9=Adam Engel} => {Batter1=Yoan Moncada} 0.1612903 0.9090909 1.523342 10



For the Nationals, perhaps the most interesting rule is that if they don’t hit any HR’s and have no stolen bases, they’re 1.58 times more likely to lose. Looking at their 2018 record being mostly wins and the support of this being 15% (good for the amount of data), this indicates that they’re mostly an offensive team, and that players with lower overall performance should be assessed for a spot on the team next season.

## lhs rhs support confidence lift count  
## [1] {Result=WIN,   
## NatsStolenBases=0} => {OpponentStolenBases=0} 0.2098765 0.8717949 1.295695 34  
## [2] {Result=WIN,   
## OpponentHRs=1} => {OpponentStolenBases=0} 0.1543210 0.8333333 1.238532 25  
## [3] {Location=HOME,   
## Result=WIN} => {OpponentStolenBases=0} 0.2037037 0.8048780 1.196241 33  
## [4] {Result=WIN,   
## DayNight=N} => {OpponentStolenBases=0} 0.2530864 0.7884615 1.171842 41  
## [5] {NatsHRs=0,   
## NatsStolenBases=0} => {Result=LOSS} 0.1543210 0.7812500 1.582031 25  
## [6] {Location=AWAY,   
## Result=WIN} => {OpponentStolenBases=0} 0.1975309 0.7804878 1.159991 32

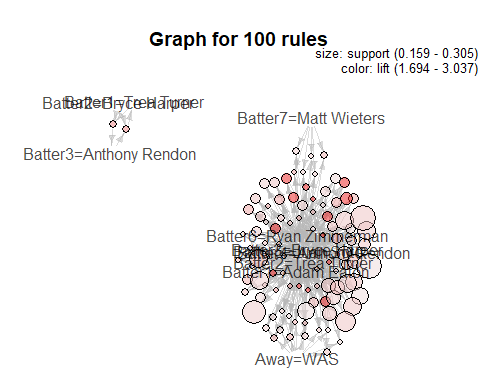
## lhs rhs support confidence lift count  
## [1] {Result=WIN,   
## NatsStolenBases=0} => {OpponentStolenBases=0} 0.2098765 0.8717949 1.2956951 34  
## [2] {Result=WIN,   
## OpponentHRs=1} => {OpponentStolenBases=0} 0.1543210 0.8333333 1.2385321 25  
## [3] {Location=HOME,   
## Result=WIN} => {OpponentStolenBases=0} 0.2037037 0.8048780 1.1962408 33  
## [4] {Result=WIN,   
## DayNight=N} => {OpponentStolenBases=0} 0.2530864 0.7884615 1.1718419 41  
## [5] {NatsHRs=0,   
## NatsStolenBases=0} => {Result=LOSS} 0.1543210 0.7812500 1.5820312 25  
## [6] {Location=AWAY,   
## Result=WIN} => {OpponentStolenBases=0} 0.1975309 0.7804878 1.1599910 32  
## [7] {NatsStolenBases=0,   
## OpponentHRs=1} => {DayNight=N} 0.1604938 0.7222222 1.1584158 26  
## [8] {NatsStolenBases=0,   
## OpponentHRs=1} => {OpponentStolenBases=0} 0.1604938 0.7222222 1.0733945 26  
## [9] {DayNight=N,   
## NatsStolenBases=0} => {OpponentStolenBases=0} 0.2654321 0.7166667 1.0651376 43  
## [10] {DayNight=N,   
## OpponentHRs=1} => {NatsStolenBases=0} 0.1604938 0.7027027 1.2648649 26  
## [11] {DayNight=N,   
## OpponentHRs=1} => {OpponentStolenBases=0} 0.1604938 0.7027027 1.0443838 26  
## [12] {Location=AWAY,   
## Result=LOSS} => {DayNight=N} 0.1728395 0.7000000 1.1227723 28  
## [13] {NatsStolenBases=0,   
## OpponentStolenBases=0} => {DayNight=N} 0.2654321 0.6935484 1.1124241 43  
## [14] {Result=WIN,   
## NatsStolenBases=0} => {DayNight=N} 0.1666667 0.6923077 1.1104341 27  
## [15] {Location=AWAY,   
## NatsStolenBases=0} => {DayNight=N} 0.1913580 0.6888889 1.1049505 31  
## [16] {Location=AWAY,   
## NatsStolenBases=0} => {OpponentStolenBases=0} 0.1913580 0.6888889 1.0238532 31  
## [17] {Location=HOME,   
## NatsStolenBases=0} => {OpponentStolenBases=0} 0.1913580 0.6888889 1.0238532 31  
## [18] {Location=HOME,   
## DayNight=N} => {OpponentStolenBases=0} 0.2037037 0.6875000 1.0217890 33  
## [19] {Location=AWAY,   
## DayNight=N} => {OpponentStolenBases=0} 0.2222222 0.6792453 1.0095205 36  
## [20] {Result=LOSS,   
## DayNight=N} => {NatsStolenBases=0} 0.2037037 0.6734694 1.2122449 33  
## [21] {Location=HOME,   
## Result=WIN} => {DayNight=N} 0.1666667 0.6585366 1.0562666 27  
## [22] {Result=LOSS,   
## NatsHRs=0} => {NatsStolenBases=0} 0.1543210 0.6578947 1.1842105 25  
## [23] {Location=HOME,   
## Result=LOSS} => {NatsStolenBases=0} 0.1604938 0.6500000 1.1700000 26  
## [24] {Result=LOSS,   
## NatsStolenBases=0} => {DayNight=N} 0.2037037 0.6470588 1.0378567 33  
## [25] {Location=HOME,   
## NatsStolenBases=0} => {DayNight=N} 0.1790123 0.6444444 1.0336634 29  
## [26] {Location=AWAY,   
## OpponentStolenBases=0} => {DayNight=N} 0.2222222 0.6428571 1.0311174 36  
## [27] {Result=LOSS,   
## OpponentStolenBases=0} => {NatsStolenBases=0} 0.1728395 0.6363636 1.1454545 28  
## [28] {Result=LOSS,   
## OpponentStolenBases=0} => {DayNight=N} 0.1728395 0.6363636 1.0207021 28  
## [29] {Result=WIN,   
## OpponentStolenBases=0} => {DayNight=N} 0.2530864 0.6307692 1.0117289 41  
## [30] {Location=AWAY,   
## Result=LOSS} => {NatsStolenBases=0} 0.1543210 0.6250000 1.1250000 25  
## [31] {DayNight=N,   
## OpponentStolenBases=0} => {NatsStolenBases=0} 0.2654321 0.6231884 1.1217391 43  
## [32] {Location=HOME,   
## OpponentStolenBases=0} => {Result=WIN} 0.2037037 0.6226415 1.2300966 33  
## [33] {Location=HOME,   
## OpponentStolenBases=0} => {DayNight=N} 0.2037037 0.6226415 0.9986923 33  
## [34] {OpponentHRs=1,   
## OpponentStolenBases=0} => {NatsStolenBases=0} 0.1604938 0.6190476 1.1142857 26  
## [35] {OpponentHRs=1,   
## OpponentStolenBases=0} => {DayNight=N} 0.1604938 0.6190476 0.9929279 26  
## [36] {Location=AWAY,   
## Result=WIN} => {DayNight=N} 0.1543210 0.6097561 0.9780246 25  
## [37] {Location=HOME,   
## DayNight=N} => {NatsStolenBases=0} 0.1790123 0.6041667 1.0875000 29  
## [38] {OpponentHRs=1,   
## OpponentStolenBases=0} => {Result=WIN} 0.1543210 0.5952381 1.1759582 25  
## [39] {DayNight=N,   
## OpponentStolenBases=0} => {Result=WIN} 0.2530864 0.5942029 1.1739130 41  
## [40] {Location=AWAY,   
## DayNight=N} => {NatsStolenBases=0} 0.1913580 0.5849057 1.0528302 31  
## [41] {Location=HOME,   
## OpponentStolenBases=0} => {NatsStolenBases=0} 0.1913580 0.5849057 1.0528302 31  
## [42] {Location=HOME,   
## NatsStolenBases=0} => {Result=LOSS} 0.1604938 0.5777778 1.1700000 26  
## [43] {Result=LOSS,   
## DayNight=N} => {Location=AWAY} 0.1728395 0.5714286 1.1428571 28  
## [44] {Result=LOSS,   
## DayNight=N} => {OpponentStolenBases=0} 0.1728395 0.5714286 0.8492792 28  
## [45] {Location=AWAY,   
## OpponentStolenBases=0} => {Result=WIN} 0.1975309 0.5714286 1.1289199 32  
## [46] {Location=HOME,   
## DayNight=N} => {Result=WIN} 0.1666667 0.5625000 1.1112805 27  
## [47] {Location=AWAY,   
## NatsStolenBases=0} => {Result=LOSS} 0.1543210 0.5555556 1.1250000 25  
## [48] {Location=AWAY,   
## OpponentStolenBases=0} => {NatsStolenBases=0} 0.1913580 0.5535714 0.9964286 31  
## [49] {DayNight=N,   
## NatsStolenBases=0} => {Result=LOSS} 0.2037037 0.5500000 1.1137500 33  
## [50] {Result=LOSS,   
## NatsStolenBases=0} => {OpponentStolenBases=0} 0.1728395 0.5490196 0.8159741 28  
## [51] {NatsStolenBases=0,   
## OpponentStolenBases=0} => {Result=WIN} 0.2098765 0.5483871 1.0833989 34  
## [52] {Location=AWAY,   
## DayNight=N} => {Result=LOSS} 0.1728395 0.5283019 1.0698113 28  
## [53] {Result=WIN,   
## OpponentStolenBases=0} => {NatsStolenBases=0} 0.2098765 0.5230769 0.9415385 34  
## [54] {DayNight=N,   
## OpponentStolenBases=0} => {Location=AWAY} 0.2222222 0.5217391 1.0434783 36  
## [55] {Result=WIN,   
## DayNight=N} => {Location=HOME} 0.1666667 0.5192308 1.0384615 27  
## [56] {Result=WIN,   
## DayNight=N} => {NatsStolenBases=0} 0.1666667 0.5192308 0.9346154 27  
## [57] {DayNight=N,   
## NatsStolenBases=0} => {Location=AWAY} 0.1913580 0.5166667 1.0333333 31  
## [58] {Result=LOSS,   
## NatsStolenBases=0} => {Location=HOME} 0.1604938 0.5098039 1.0196078 26  
## [59] {Result=WIN,   
## OpponentStolenBases=0} => {Location=HOME} 0.2037037 0.5076923 1.0153846 33  
## [60] {NatsStolenBases=0,   
## OpponentStolenBases=0} => {Location=AWAY} 0.1913580 0.5000000 1.0000000 31  
## [61] {NatsStolenBases=0,   
## OpponentStolenBases=0} => {Location=HOME} 0.1913580 0.5000000 1.0000000 31



Trea Turner was in 12 of the top 20 rules for the winning batting lineups at either first or second in the lineup, indicating that a strong leadoff is necessary for a win. Juan Soto made a difference when they brought him up from the minors, with 8 of the top 20 rules having him batting at fifth. Like with the White Sox, the remaining players’ performance should be assessed for a remaining spot on the team.

## lhs rhs support confidence lift count  
## [1] {Batter2=Bryce Harper,   
## Batter3=Anthony Rendon} => {Batter1=Trea Turner} 0.1707317 1 2.484848 14  
## [2] {Batter1=Adam Eaton,   
## Batter6=Ryan Zimmerman} => {Batter4=Anthony Rendon} 0.1707317 1 3.037037 14  
## [3] {Batter5=Juan Soto,   
## Batter6=Ryan Zimmerman} => {Batter4=Anthony Rendon} 0.1951220 1 3.037037 16  
## [4] {Batter4=Anthony Rendon,   
## Batter6=Ryan Zimmerman} => {Batter2=Trea Turner} 0.2073171 1 1.952381 17  
## [5] {Batter2=Trea Turner,   
## Batter6=Ryan Zimmerman} => {Batter4=Anthony Rendon} 0.2073171 1 3.037037 17  
## [6] {Batter3=Bryce Harper,   
## Batter6=Ryan Zimmerman} => {Batter4=Anthony Rendon} 0.1951220 1 3.037037 16  
## [7] {Batter1=Adam Eaton,   
## Batter6=Ryan Zimmerman} => {Batter2=Trea Turner} 0.1707317 1 1.952381 14  
## [8] {Batter5=Juan Soto,   
## Batter6=Ryan Zimmerman} => {Batter2=Trea Turner} 0.1951220 1 1.952381 16  
## [9] {Batter5=Juan Soto,   
## Batter6=Ryan Zimmerman} => {Batter3=Bryce Harper} 0.1951220 1 1.863636 16  
## [10] {Batter3=Bryce Harper,   
## Batter6=Ryan Zimmerman} => {Batter5=Juan Soto} 0.1951220 1 2.277778 16  
## [11] {Batter3=Bryce Harper,   
## Batter6=Ryan Zimmerman} => {Batter2=Trea Turner} 0.1951220 1 1.952381 16  
## [12] {Batter4=Anthony Rendon,   
## Batter7=Matt Wieters} => {Batter2=Trea Turner} 0.1585366 1 1.952381 13  
## [13] {Batter5=Juan Soto,   
## Batter7=Matt Wieters} => {Batter3=Bryce Harper} 0.1707317 1 1.863636 14  
## [14] {Batter1=Adam Eaton,   
## Batter4=Anthony Rendon} => {Batter2=Trea Turner} 0.2682927 1 1.952381 22  
## [15] {Batter4=Anthony Rendon,   
## Batter5=Juan Soto} => {Batter2=Trea Turner} 0.2560976 1 1.952381 21  
## [16] {Batter4=Anthony Rendon,   
## Batter5=Juan Soto} => {Batter3=Bryce Harper} 0.2560976 1 1.863636 21  
## [17] {Home=WAS,   
## Batter4=Anthony Rendon} => {Batter2=Trea Turner} 0.1585366 1 1.952381 13  
## [18] {Away=WAS,   
## Batter4=Anthony Rendon} => {Batter2=Trea Turner} 0.1707317 1 1.952381 14  
## [19] {Batter3=Bryce Harper,   
## Batter4=Anthony Rendon} => {Batter2=Trea Turner} 0.3048780 1 1.952381 25  
## [20] {Batter1=Adam Eaton,   
## Batter5=Juan Soto,   
## Batter6=Ryan Zimmerman} => {Batter4=Anthony Rendon} 0.1585366 1 3.037037 13

## Warning: plot: Too many rules supplied. Only plotting the best 100 rules using  
## 'support' (change control parameter max if needed)



## Model 2

Each team and each position results, batters versus pitchers, were added to a confusion matrix. For the Chicago White Sox batters, all three high salary players were classified as low salaried. One low salaried player was ranked as high salaried while three were classified as mid salaried. The remaining six were correctly classified. All three mid salaried players were ranked as low salaried. At 37.5% accurate, this model is not accurate as many low paid players play as well or better than the higher paid players.

## [1] "Chicago White Sox Batter Salary Prediction Accuracy"

##   
## high low mid  
## high 0 3 0  
## low 2 6 2  
## mid 0 3 0

## [1] 37.5

For the Chicago White Sox pitchers, both high salary players were classified as low salaried. Two low salaried players were ranked as mid salaried while the remaining six were correctly classified. One mid salaried pitcher was correctly classified and the other four were classified as low salaried. At 46.67% accurate, this model is not accurate either. A case can be made that the White Sox were undervaluing several players and overvaluing others.

## [1] "Chicago White Sox Pitcher Salary Prediction Accuracy"

##   
## high low mid  
## high 0 2 0  
## low 0 7 1  
## mid 0 4 1

## [1] 53.33333

The Washington Nationals batters have a more mixed prediction with lower accuracy at just 21.05%. For the Nationals, there is a better mix between low, mid, and high salaried players where the White Sox, for the most part, either paid their players close to the league minimum, 545,000, or a very high salary, 7 million or more. Only four players were accurately predicted for salary, 2 high and 2 low. Three high paid players were ranked as low while two were raked as mid. Of the misclassified low paid players, two were ranked as high and three were ranked as mid. The mid salaried players were mostly classified as high with one classified as low.

## [1] "Washington Nationals Batter Salary Prediction Accuracy"

##   
## high low mid  
## high 0 5 2  
## low 3 2 2  
## mid 2 2 1

## [1] 15.78947

The best prediction accuracy was the pitching staff of the Washington Nationals at 53.33%. Two of the four high paid pitchers were misclassified as mid salary while the other two were accurately predicted. Two of the seven low salaried pitchers were misclassified as mid salary. The other five low salaried pitchers were accurately predicted. For the mid salary group, two of four were classified as low. Of the remaining two pitchers, one was correctly classified as mid-range. Unlike the White Sox, the Nationals are not overwhelmingly under- or over-valuing their pitching staff. However, for batters it appears that the Nationals are consistent in mis-valuing their players.

## [1] "Washington Nationals Pitcher Salary Prediction Accuracy"

##   
## high low mid  
## high 3 0 1  
## low 1 5 1  
## mid 1 3 0

## [1] 53.33333

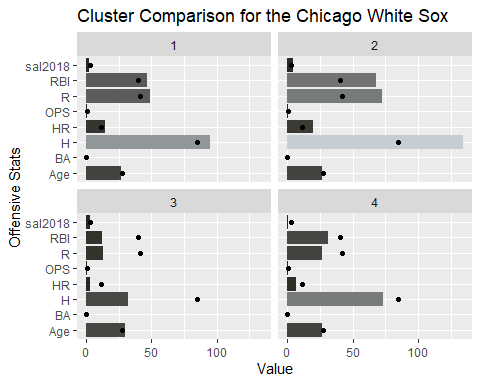
## Model 3

When k = 2, the clusters are split based where cluster 1 has scored greater than 45 runs, and cluster 2 has scored less than 45 runs. There is slight overlap with 3 players (Adam Engel, Leury Garcia, Omar Narvaez) in terms of separating them into clusters based on RBIs. Overall, cluster two offensively performs better than cluster 1.

K=4 provides a more detailed breakdown of the runs scored by the White Sox players, where the players can be visually grouped into 4 offensive tiers: High Performing Tier, Upper Middle Performing Tier, Lower Middle Performing Tier, Lower Performing Tier.

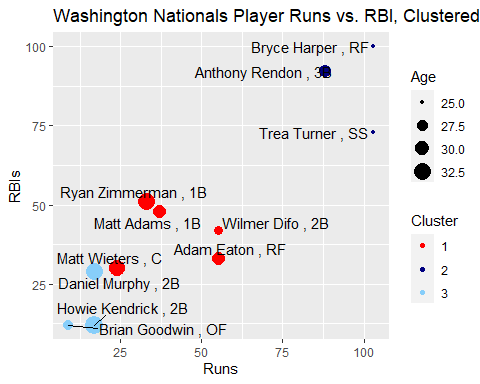
K=6 follows upon the 4-means clustering above and further separates players based on offensive performance. Tim Anderson and Yoan Moncada form the new cluster 5, while Avisail Garcia and Adam Engel form the new cluster 6. Matt Davidson remain. With 6 centroids, Matt Davidson and Jose Abreu form the first cluster, whereas in earlier k-means models (k=4) Matt Davidson was not with Jose Abreu. Both Matt Davidson and Jose Abreu had higher OPS and OBP compared to Tim Anderson and Yoan Moncada.

K=4 did not have the smallest sum of squared errors (as shown in the elbow plot) but showed the best separation between high performing and low performing White Sox players. It also provided the best business context sense that is easy to interpret.

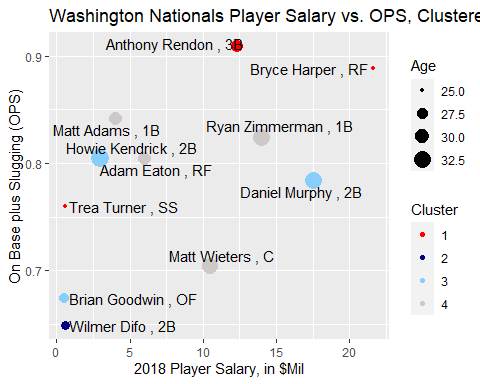


## Player Player2 X1 X2 Position MLSRV  
## 3 Jose Abreu Abreu, Jose Abreu Jose 1b 4.000  
## 10 Welington Castillo Castillo, Welington Castillo Welington c 6.009  
## 2 Avisail Garcia Garcia, Avisail Garcia Avisail rf 4.167  
## 5 Leury Garcia Garcia, Leury Garcia Leury cf-2b 3.025  
## 8 Tim Anderson Anderson, Tim Anderson Tim ss 1.115  
## 6 Matt Davidson Davidson, Matt Davidson Matt dh-3b 1.145  
## 9 Tyler Saladino Saladino, Tyler Saladino Tyler 2b 2.087  
## 4 Kevan Smith Smith, Kevan Smith Kevan c 1.043  
## 7 Omar Narvaez Narvaez, Omar Narvaez Omar c 1.089  
## 11 Yoan Moncada Moncada, Yoan Moncada Yoan 2b-3b 0.106  
## 1 Adam Engel Engel, Adam Engel Adam of 0.118  
## Agent Length.TotalValue sal2018 Rk Pos Age G PA AB  
## 3 ISE Baseball 1 yr/$13M (18) 13.000 2 1B 31 128 553 499  
## 10 ACES 2 yr/$15M (18-19)+20 cl opt 7.250 13 C 31 49 181 170  
## 2 Gene Mato 1 yr/$6.7M (18) 6.700 8 RF 27 93 385 356  
## 5 Rep 1 Baseball 1 yr/$1.175M (18) 1.175 11 OF 27 82 275 258  
## 8 Reynolds Sports 6 yr/$25M (17-22)+23-24 opts 1.000 4 SS 25 153 606 567  
## 6 MVP Sports 1 yr/$0.57M (18) 0.570 9 DH 27 123 496 434  
## 9 1 yr/$0.565M (18) 0.565 21 DH 28 6 9 8  
## 4 Pro Star Mgt 1 yr/$0.56M (18) 0.560 12 C 30 52 187 171  
## 7 1 yr/$0.56M (18) 0.560 1 C 26 97 322 280  
## 11 David Hastings 1 yr/$0.555M (18) 0.555 3 2B 23 149 650 578  
## 1 1 yr/$0.552M (18) 0.552 7 CF 26 143 463 429  
## R H X2B X3B HR RBI SB CS BB SO BA OBP SLG OPS OPS. TB GDP HBP  
## 3 68 132 36 1 22 78 2 0 37 109 0.265 0.325 0.473 0.798 117 236 14 11  
## 10 17 44 7 0 6 15 1 0 9 46 0.259 0.304 0.406 0.710 94 69 7 2  
## 2 47 84 11 2 19 49 3 1 20 102 0.236 0.281 0.438 0.719 95 156 9 4  
## 5 23 70 7 4 4 32 12 1 9 69 0.271 0.303 0.376 0.679 87 97 2 3  
## 8 77 136 28 3 20 64 26 8 30 149 0.240 0.281 0.406 0.687 87 230 15 4  
## 6 51 99 23 0 20 62 0 0 52 165 0.228 0.319 0.419 0.738 102 182 8 7  
## 9 2 2 1 0 0 0 0 0 0 3 0.250 0.250 0.375 0.625 70 3 0 0  
## 4 21 50 6 0 3 21 1 0 10 18 0.292 0.348 0.380 0.728 102 65 5 5  
## 7 30 77 14 1 9 30 0 2 38 65 0.275 0.366 0.429 0.794 119 120 5 2  
## 11 73 136 32 6 17 61 12 6 67 217 0.235 0.315 0.400 0.714 96 231 4 1  
## 1 49 101 17 4 6 29 16 8 18 129 0.235 0.279 0.336 0.614 69 144 1 8  
## SH SF IBB BATS salCategory cws\_four\_clusters.cluster  
## 3 0 6 7 RIGHT high 2  
## 10 0 0 0 RIGHT high 3  
## 2 0 5 2 RIGHT mid 1  
## 5 4 1 0 SWITCH low 4  
## 8 2 3 2 RIGHT low 2  
## 6 0 3 0 RIGHT low 1  
## 9 1 0 0 RIGHT low 3  
## 4 0 1 0 RIGHT low 3  
## 7 2 0 1 LEFT low 4  
## 11 2 2 1 SWITCH low 2  
## 1 7 1 0 RIGHT low 1

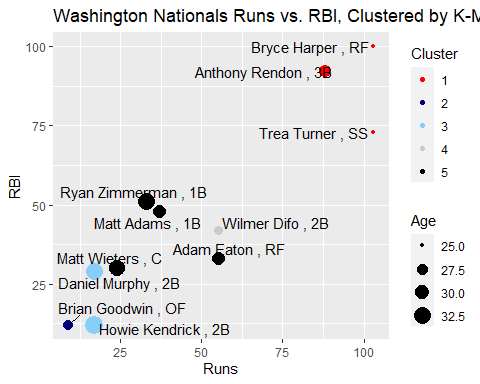
K=3 is given as the optimal k value via the Elbow Method, and clear separation between clusters is seen when viewing RBIs against Runs. The three players with the most RBIs, Turner, Rendon, and Harper, are all members of the 2nd cluster, while the 3rd cluster contains the players with the least RBIs.



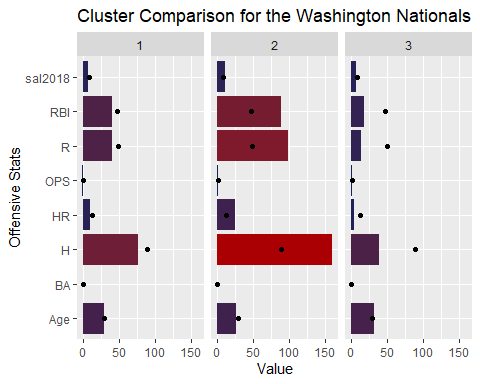
When the kmeans algorithm uses 4 clusters, Wilmer Difo is separated from cluster 1 and forms his own new cluster. However, he still has similar Runs and RBIs to Matt Adams and Adam Eaton, who are in cluster two. At this k value, we can begin to see the individual examples become their own clusters, as k approaches the number of samples.



With 5 clusters, Wilmer Difo is still isolated in a separate cluster, and Daniel Murphy and Howie Kendrick both leave cluster 3 to form their own. Brian Goodwin also forms a solitary cluster as well. The k-means algorithm continues to create single unit clusters as k approaches the number of samples.



K = 3 provided the best delineation between players, although 4 and 5 had lower Within cluster sum of squared errors. As k approaches n, the number of samples in the dataset, individual players will begin forming solitary clusters. Ultimately, when k = n, each player will have his own cluster. This type of clustering is analogous to overfitting in other predictive models.



## Player Player2 X1 X2 Position MLSRV  
## 4 Bryce Harper Harper, Bryce Harper Bryce rf 5.159  
## 5 Daniel Murphy Murphy, Daniel Murphy Daniel 2b 8.109  
## 9 Ryan Zimmerman Zimmerman, Ryan Zimmerman Ryan 1b 12.032  
## 2 Anthony Rendon Rendon, Anthony Rendon Anthony 3b 4.130  
## 8 Matt Wieters Wieters, Matt Wieters Matt c 8.129  
## 1 Adam Eaton Eaton, Adam Eaton Adam cf 5.030  
## 7 Matt Adams Adams, Matt Adams Matt 1b 5.033  
## 6 Howie Kendrick Kendrick, Howie Kendrick Howie lf 11.091  
## 10 Trea Turner Turner, Trea Turner Trea ss 1.135  
## 11 Wilmer Difo Difo, Wilmer Difo Wilmer ss 1.110  
## 3 Brian Goodwin Goodwin, Brian Goodwin Brian of 1.019  
## Agent Length.TotalValue sal2018 Rk Pos Age G PA  
## 4 Boras Corp. 1 yr/$21.625M (18) 21.6250 8 RF 25 159 695  
## 5 ACES 3 yr/$37.5M (16-18) 17.5000 13 2B 33 56 205  
## 9 CAA Sports 6 yr/$100M (14-19)+20 opt 14.0000 2 1B 33 85 323  
## 2 Boras Corp. 1 yr/$12.3M (18) 12.3000 5 3B 28 136 597  
## 8 Boras Corp. 1 yr/$10.5M (17)+18 p opt 10.5000 1 C 32 76 271  
## 1 Diamond Spts 5 yr/$23.5M (15-19)+20-21 opts 6.0000 9 RF 29 95 370  
## 7 Wasserman Media 1 yr/$4M (18) 4.0000 10 1B 29 94 277  
## 6 Reynolds Sports 2 yr/$7M (18-19) 3.0000 14 2B 34 40 160  
## 10 CAA Sports 1 yr/$0.5772M (18) 0.5772 4 SS 25 162 740  
## 11 1 yr/$0.5579M (18) 0.5579 3 2B 26 148 456  
## 3 Boras Corp. 1 yr/$0.5539M (18) 0.5539 17 OF 27 48 79  
## AB R H X2B X3B HR RBI SB CS BB SO BA OBP SLG OPS OPS. TB  
## 4 550 103 137 34 0 34 100 13 3 130 169 0.249 0.393 0.496 0.889 133 273  
## 5 190 17 57 9 0 6 29 1 0 13 17 0.300 0.341 0.442 0.784 105 84  
## 9 288 33 76 21 2 13 51 1 1 30 55 0.264 0.337 0.486 0.824 114 140  
## 2 529 88 163 44 2 24 92 2 1 55 82 0.308 0.374 0.535 0.909 137 283  
## 8 235 24 56 8 0 8 30 0 1 30 45 0.238 0.330 0.374 0.704 86 88  
## 1 319 55 96 18 1 5 33 9 1 38 64 0.301 0.394 0.411 0.805 114 131  
## 7 249 37 64 9 0 18 48 0 0 24 55 0.257 0.332 0.510 0.842 118 127  
## 6 152 17 46 14 0 4 12 1 1 5 29 0.303 0.331 0.474 0.805 110 72  
## 10 664 103 180 27 6 19 73 43 9 69 132 0.271 0.344 0.416 0.760 100 276  
## 11 408 55 94 14 7 7 42 10 3 39 82 0.230 0.298 0.350 0.649 71 143  
## 3 65 9 13 1 0 3 12 3 1 10 26 0.200 0.321 0.354 0.674 78 23  
## GDP HBP SH SF IBB BATS salCategory three\_clusters.cluster  
## 4 7 6 0 9 16 LEFT high 2  
## 5 4 0 0 2 2 LEFT high 3  
## 9 10 3 0 2 1 RIGHT high 1  
## 2 5 5 0 8 5 RIGHT high 2  
## 8 5 3 1 2 3 SWITCH mid 1  
## 1 2 11 2 0 0 LEFT mid 1  
## 7 6 4 0 0 2 LEFT mid 1  
## 6 6 2 0 1 1 RIGHT low 3  
## 10 7 5 2 0 3 RIGHT low 2  
## 11 8 2 3 4 5 SWITCH low 1  
## 3 0 2 1 1 0 LEFT low 3

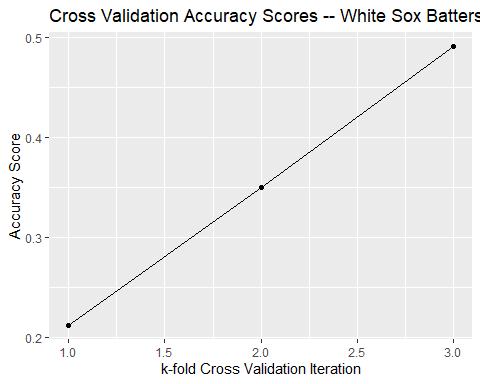
The second Cluster was the highest performing for all offensive categories. The greatest difference is observed in the sum total stats, rather than the calculated stats. (i.e., Hits and Runs rather than OPS and BA). This suggest that a general manager should aim to use the pure stats to review player performance and make comparisons rather than the calculated stats (assuming equal number of games played).

## Model 4

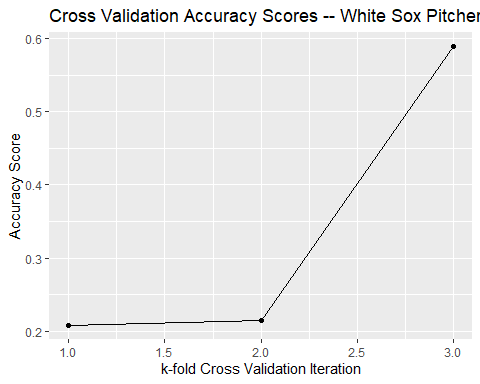
True accuracy scores were measured to test model validity. True accuracy is represented as: (Count(Total Correct Predictions))/(Count(All Predictions))

The 3-fold CV resulted in an average accuracy of 36.1% for the Chicago White Sox batters. For their pitchers, the 3-fold CV resulted in an average accuracy of 46.7%.

## [1] "Average Accuracy: 0.351"

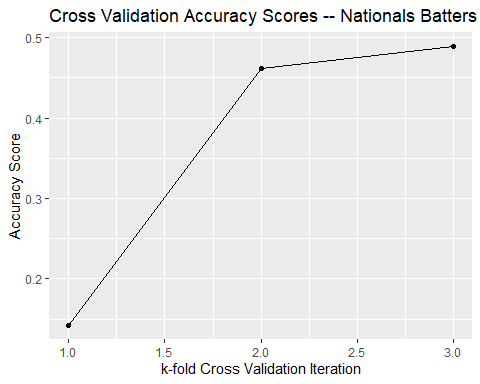


## [1] "Average Accuracy: 0.337333333333333"

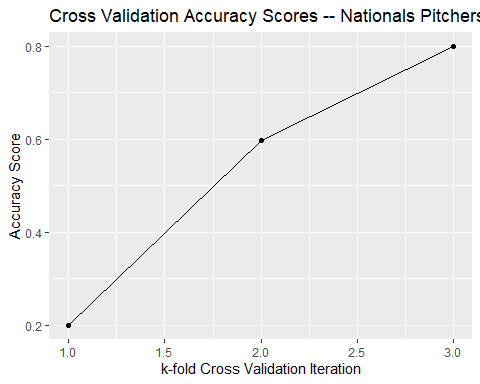


The 3-fold CV resulted in an average accuracy of 26.1% for the Washington Nationals batters. The 3-fold CV resulted in an average accuracy of 45.2% for their pitchers.

## [1] "Average Accuracy: 0.364333333333333"



## [1] "Average Accuracy: 0.532333333333333"



# Conclusions

Can machine learning algorithms identify who they should trade in order to play a little Moneyball? The answer is a qualified yes. The algorithms definitely help to identify players whose salary does not match other players with similar records. However, even with the identification, there needs to be human interpretation of the results.

The Chicago White Sox pay Jose Abreu handsomely, but random forest assigned him to the low salary group. Why? It is not because he is not a good player as his offensive power statistic (OPS), RBIs, and slugging (SLG) are the highest on the team and he is among the highest for on-base percentage (OBP) and batting average. What is discovered through K-Means clustering is that many of the low paid players are similar to Abreu.

Based on statistics and identification by the different models, the team would be best served by trading the following position players: Adam Engel and Welington Castillo. Castillo is the clearest case as he is high salary but made few appearances with few RBIs and a mediocre batting average. Adam Engel is not on the mismatched salary list however, his batting average amongst the lowest on the team when looking at players with more than 150 at bats. He also has low OBP, SLG, and OPS.

## Player2 orig pred Rk AB RBI BA OBP SLG OPS  
## 1 Abreu, Jose high low 2 499 78 0.265 0.325 0.473 0.798  
## 2 Avilan, Luis mid low 28 1 0 0.000 0.000 0.000 0.000  
## 3 Castillo, Welington high low 13 170 15 0.259 0.304 0.406 0.710  
## 4 Garcia, Avisail mid low 8 356 49 0.236 0.281 0.438 0.719  
## 5 Giolito, Lucas low mid 24 6 0 0.000 0.000 0.000 0.000  
## 6 Lopez, Reynaldo low mid 29 1 0 0.000 0.000 0.000 0.000  
## 7 Saladino, Tyler low high 21 8 0 0.250 0.250 0.375 0.625  
## 8 Santiago, Hector mid low 25 4 0 0.000 0.000 0.000 0.000  
## 9 Shields, James high low 23 6 0 0.167 0.167 0.167 0.333  
## 10 Smith, Kevan low high 12 171 21 0.292 0.348 0.380 0.728

## Player2 orig pred Rk Pos AB R RBI BA OBP SLG OPS  
## 1 Anderson, Tim low low 4 SS 567 77 64 0.240 0.281 0.406 0.687  
## 2 Davidson, Matt low low 9 DH 434 51 62 0.228 0.319 0.419 0.738  
## 3 Engel, Adam low low 7 CF 429 49 29 0.235 0.279 0.336 0.614  
## 4 Garcia, Leury low low 11 OF 258 23 32 0.271 0.303 0.376 0.679  
## 5 Moncada, Yoan low low 3 2B 578 73 61 0.235 0.315 0.400 0.714  
## 6 Narvaez, Omar low low 1 C 280 30 30 0.275 0.366 0.429 0.794

For the Chicago pitchers, the trade recommendations are more obvious. Both earned run average ERA and fielding independent percentage (FIP) are the best metrics for pitchers and should be low. Miguel Gonzalez has the highest ERA and FIP. Even more glaring for Gonzalez, he has won no games. Both and Carlos Rodon are in the mid-salary range but do not have the statistics to support such pay. Rodon has a losing record, an ERA over the league average of 4.15, and a FIP of nearly 5. Both Carson Fulmer and Lucas Giolito are low paid pitchers but it is still recommended to trade them as they have a losing record with high ERAs and high FIPS.

## Player2 orig pred Rk W.L. ERA SV FIP  
## 1 Avilan, Luis mid low 10 0.667 3.86 2 2.71  
## 2 Bummer, Aaron low mid 13 0.000 4.26 0 2.40  
## 3 Gonzalez, Miguel mid low 21 0.000 12.41 0 8.02  
## 4 Jones, Nate mid low 14 0.500 3.00 5 4.56  
## 5 Rodon, Carlos mid low 5 0.429 4.18 0 4.95  
## 6 Shields, James high low 1 0.304 4.53 0 5.09  
## 7 Soria, Joakim high low 6 0.000 2.56 16 2.15

## Player2 orig pred Rk W.L. ERA SV FIP  
## 1 Davidson, Matt low low 30 0.000 0.00 0 2.83  
## 2 Farquhar, Danny low low 25 0.500 5.63 0 5.79  
## 3 Fulmer, Carson low low 12 0.333 8.07 0 7.27  
## 4 Giolito, Lucas low low 3 0.435 6.13 0 5.56  
## 5 Infante, Gregory low low 23 0.500 8.00 0 4.49  
## 6 Lopez, Reynaldo low low 2 0.412 3.91 0 4.63  
## 7 Minaya, Juan low low 9 0.500 3.28 1 3.57  
## 8 Santiago, Hector mid mid 7 0.667 4.41 2 5.09

The Washington Nationals should look into trading Wilmer Difo, Brian Goodwin, and Matt Wieters. Of these, only Matt Wieters is paid over $2 million dollars. Brian Goodwin has the lowest batting average of any position player and the second lowest OBP, SLG, and OPS of any position player. Difo holds the lowest OBP, SLG, and OPS of the position players and the second worst batting average. Wieters is the third worst position player for all four of those statistics.

## Player2 orig pred Rk AB RBI BA OBP SLG OPS  
## 1 Adams, Matt mid high 10 249 48 0.257 0.332 0.510 0.842  
## 2 Difo, Wilmer low mid 3 408 42 0.230 0.298 0.350 0.649  
## 3 Eaton, Adam mid low 9 319 33 0.301 0.394 0.411 0.805  
## 4 Gonzalez, Gio high low 27 44 0 0.068 0.068 0.091 0.159  
## 5 Goodwin, Brian low high 17 65 12 0.200 0.321 0.354 0.674  
## 6 Harper, Bryce high low 8 550 100 0.249 0.393 0.496 0.889  
## 7 Kelley, Shawn mid low 38 1 0 0.000 0.000 0.000 0.000  
## 8 Kendrick, Howie low high 14 152 12 0.303 0.331 0.474 0.805  
## 9 Murphy, Daniel high low 13 190 29 0.300 0.341 0.442 0.784  
## 10 Rendon, Anthony high low 5 529 92 0.308 0.374 0.535 0.909  
## 11 Roark, Tanner mid high 25 58 8 0.190 0.217 0.259 0.475  
## 12 Ross, Joe low mid 32 5 0 0.000 0.000 0.000 0.000  
## 13 Scherzer, Max high mid 24 70 6 0.243 0.274 0.271 0.545  
## 14 Strasburg, Stephen high low 26 41 1 0.122 0.163 0.122 0.285  
## 15 Turner, Trea low high 4 664 73 0.271 0.344 0.416 0.760  
## 16 Zimmerman, Ryan high mid 2 288 51 0.264 0.337 0.486 0.824

## Player2 orig pred Rk AB RBI BA OBP SLG OPS  
## 1 Cole, A.J. low low 33 3 1 0.333 0.333 1.333 1.667  
## 2 Grace, Matt low low 35 3 0 0.333 0.333 0.333 0.667  
## 3 Wieters, Matt mid mid 1 235 30 0.238 0.330 0.374 0.704

With the Washington Nationals pitching staff, Ryan Madson, Tanner Roark, Sammy Solis, AJ Cole, Gio Gonzalez, Trevor Gott, and Enny Romero should all be under consideration for trades. Madson, who earns a high salary, has four saves which is the second highest on the team but, as a reliever, he is the losing pitcher almost as often as he either wins or gets a save. His ERA and FIP are also above the league average. Roark is a mid-salary pitcher, but he is often credited with the loss. Additionally, his ERA and FIP are slightly higher than the league average. Solis has the third highest ERA and a high FIP. His losing record does not show any saves. Cole and Romero duked it out to see who could get the highest ERA and highest FIP. In both cases, Romero “won”. With ERAs over 13, both Cole and Romero should be looking for new teams. Gio Gonzalez, despite having an ERA that is only slightly above the league average, had an incredibly bad win-loss record. As he earns a high salary, it is necessary for the team to look for a pitcher that can bring the wins at a lower cost. Gott has no wins, no saves, and an ERA above the league average. It may have been a fluke season but there is nothing to recommend him to not be traded.

## Player2 orig pred Rk W.L. ERA SV FIP  
## 1 Doolittle, Sean mid low 7 0.500 1.60 25 1.89  
## 2 Grace, Matt low mid 8 0.500 2.87 0 3.39  
## 3 Kelley, Shawn mid low 15 1.000 3.34 0 4.55  
## 4 Kintzler, Brandon mid low 14 0.333 3.59 2 3.44  
## 5 Madson, Ryan high mid 10 0.286 5.28 4 4.36  
## 6 Roark, Tanner mid high 2 0.375 4.34 0 4.27  
## 7 Solis, Sammy low high 11 0.333 6.41 0 4.91

## Player2 orig pred Rk W.L. ERA SV FIP  
## 1 Cole, A.J. low low 25 0.500 13.06 0 10.51  
## 2 Glover, Koda low low 22 0.250 3.31 1 4.69  
## 3 Gonzalez, Gio high high 3 0.389 4.57 0 4.25  
## 4 Gott, Trevor low low 19 0.000 5.68 0 6.21  
## 5 Romero, Enny low low 29 0.000 13.50 0 10.66  
## 6 Ross, Joe low low 23 0.000 5.06 0 5.85  
## 7 Scherzer, Max high high 1 0.720 2.53 0 2.65  
## 8 Strasburg, Stephen high high 4 0.588 3.74 0 3.62

Of all the position players trade recommendations for either team, only Brian Goodwin and Matt Wieters appear to have been traded between the 2018 and 2019 seasons. The pitcher recommendations were followed by the Washington Nationals to the man. The Chicago White Sox did not follow any of the recommendations. It is not hard to see why the Nationals won the World Series in 2019 while the White Sox continued to lose more than they won.