

# NBA Shots Analysis

Michael Armesto, David Doman, N'Dea Jackson, Matthew McDonnell

# Brief Background

- Sourced from Kaggle
- 122,106 rows & 30 columns
- Situational variables (dribbles, distance from hoop) and scenario variables (opponent, venue location)
- Some data cleaning was performed such as changing variable types and discretizing certain variables
- Some random sampling was used to perform more efficient SVM analysis



# Clustering- Player Types

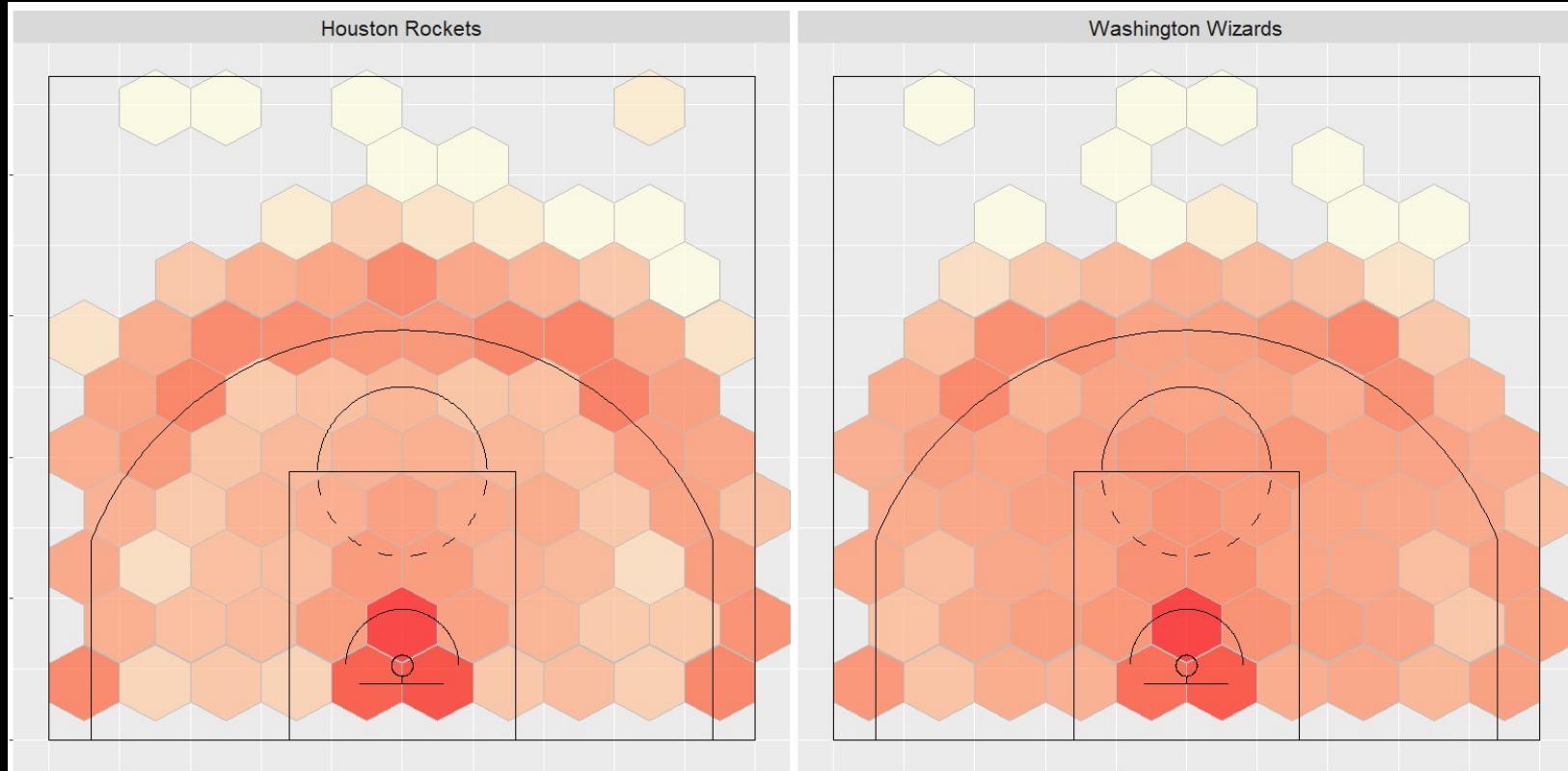
Clustered players by offensive style  
based on shot selection

Shots broken into 5 zones:

1. Above the Break 3
2. Corner 3
3. Midrange
4. Restricted Area
5. In the Paint (non-RA)

Variables used:

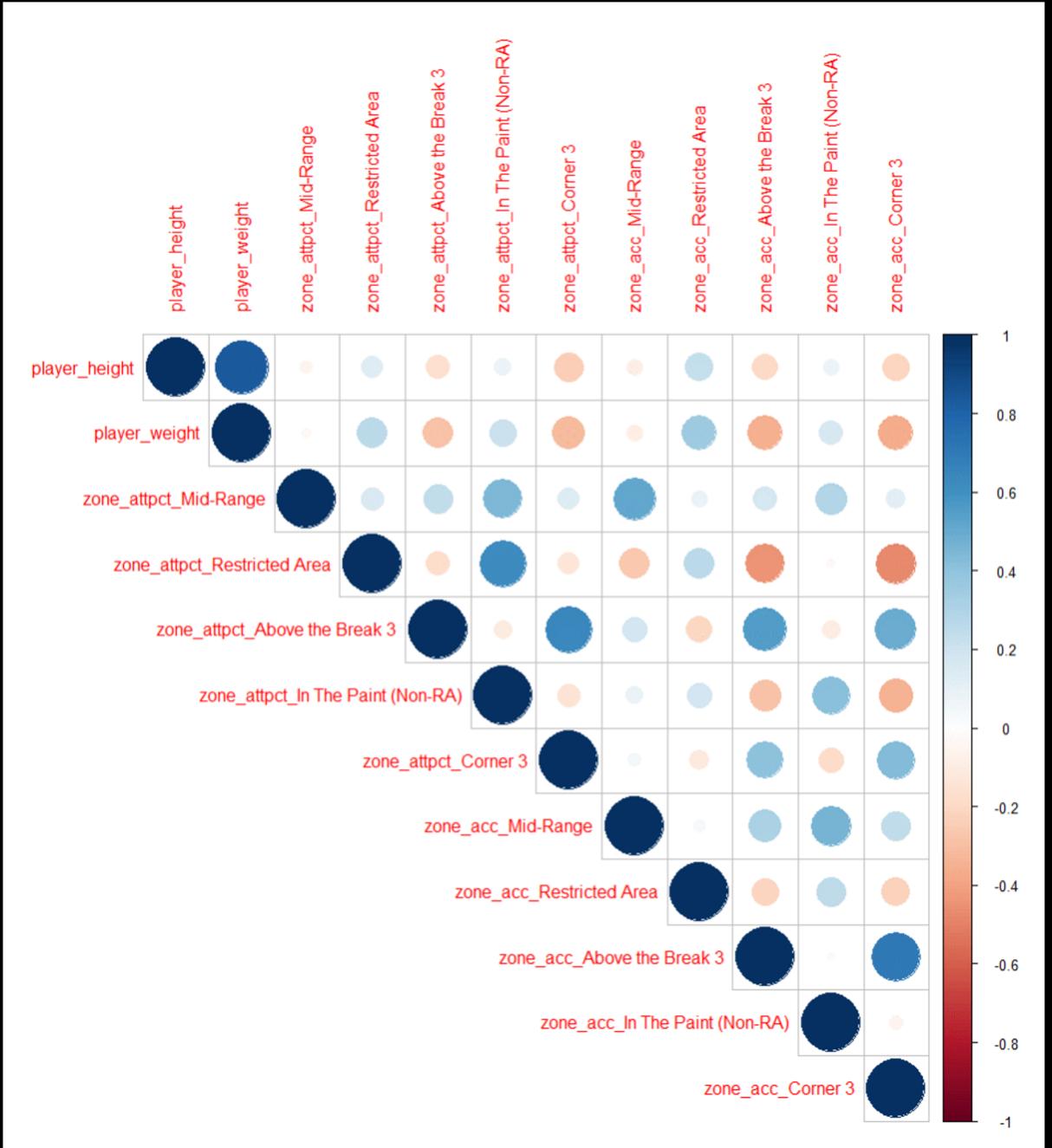
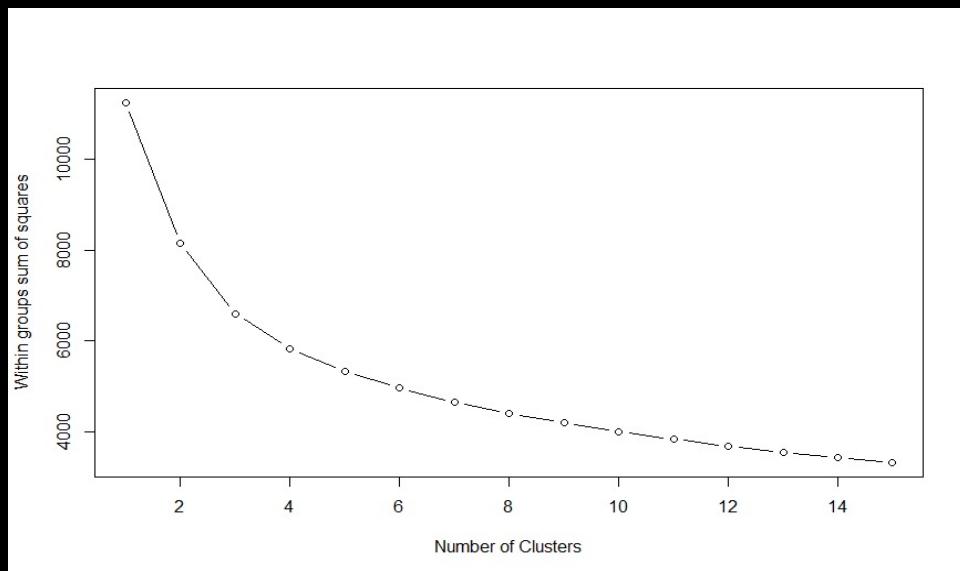
- Attempt percentage in each zone
- Relative accuracy in each zone



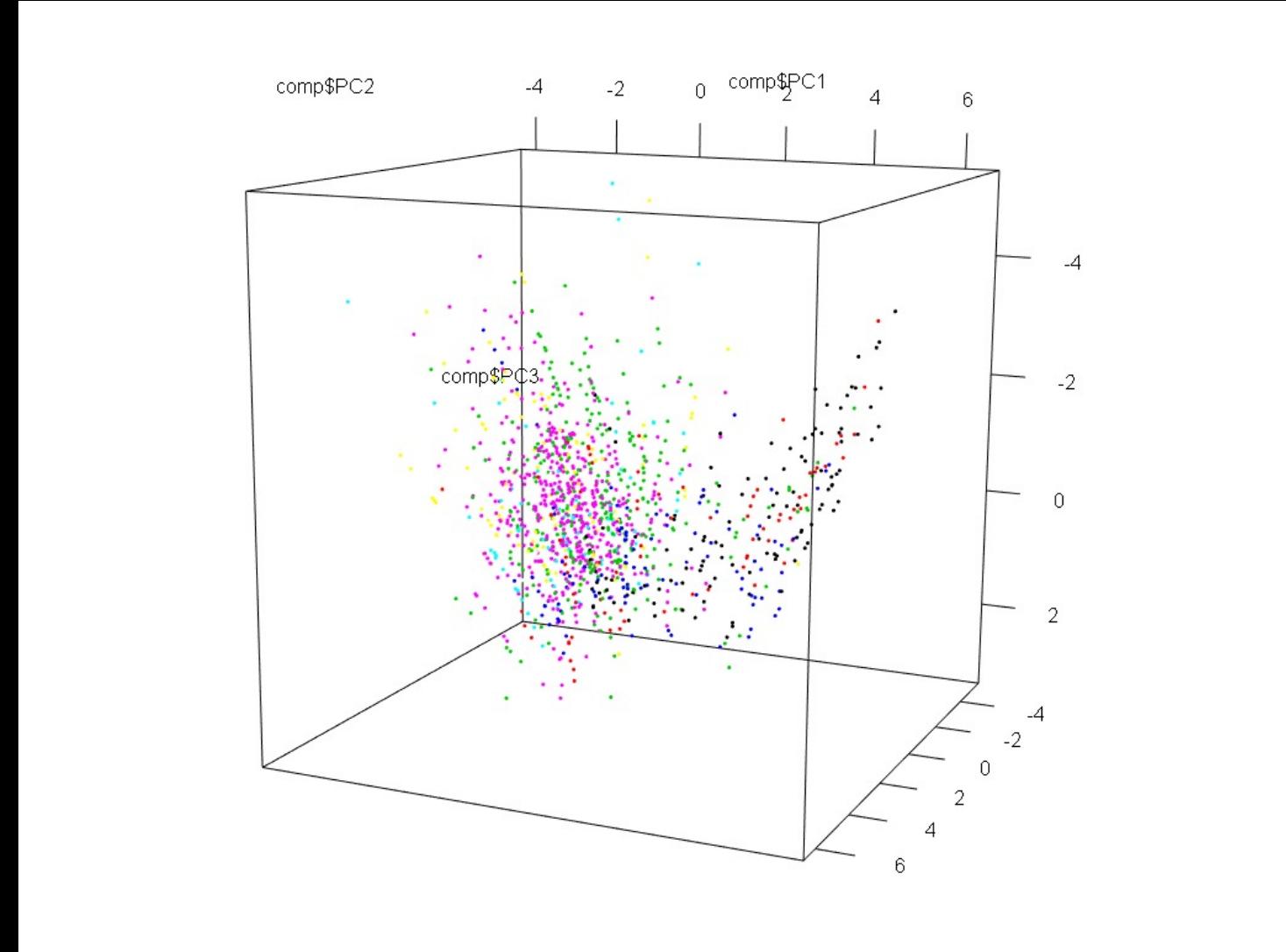
# Clustering

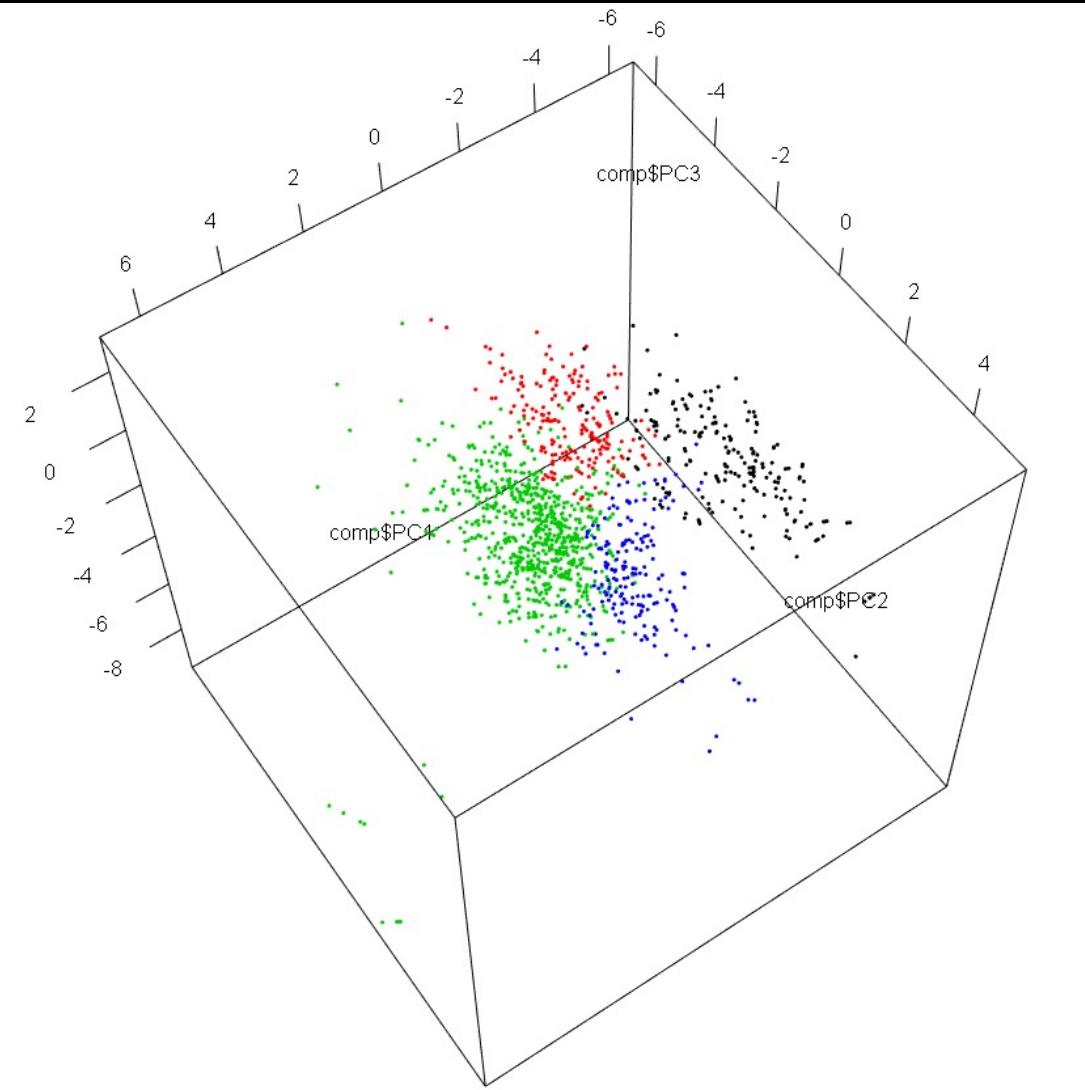
Principal Component Analysis was used due to high correlation in variables

Most impactful variables on PCA were shot attempt percentages by zone, particularly corner and above-break three-pointers

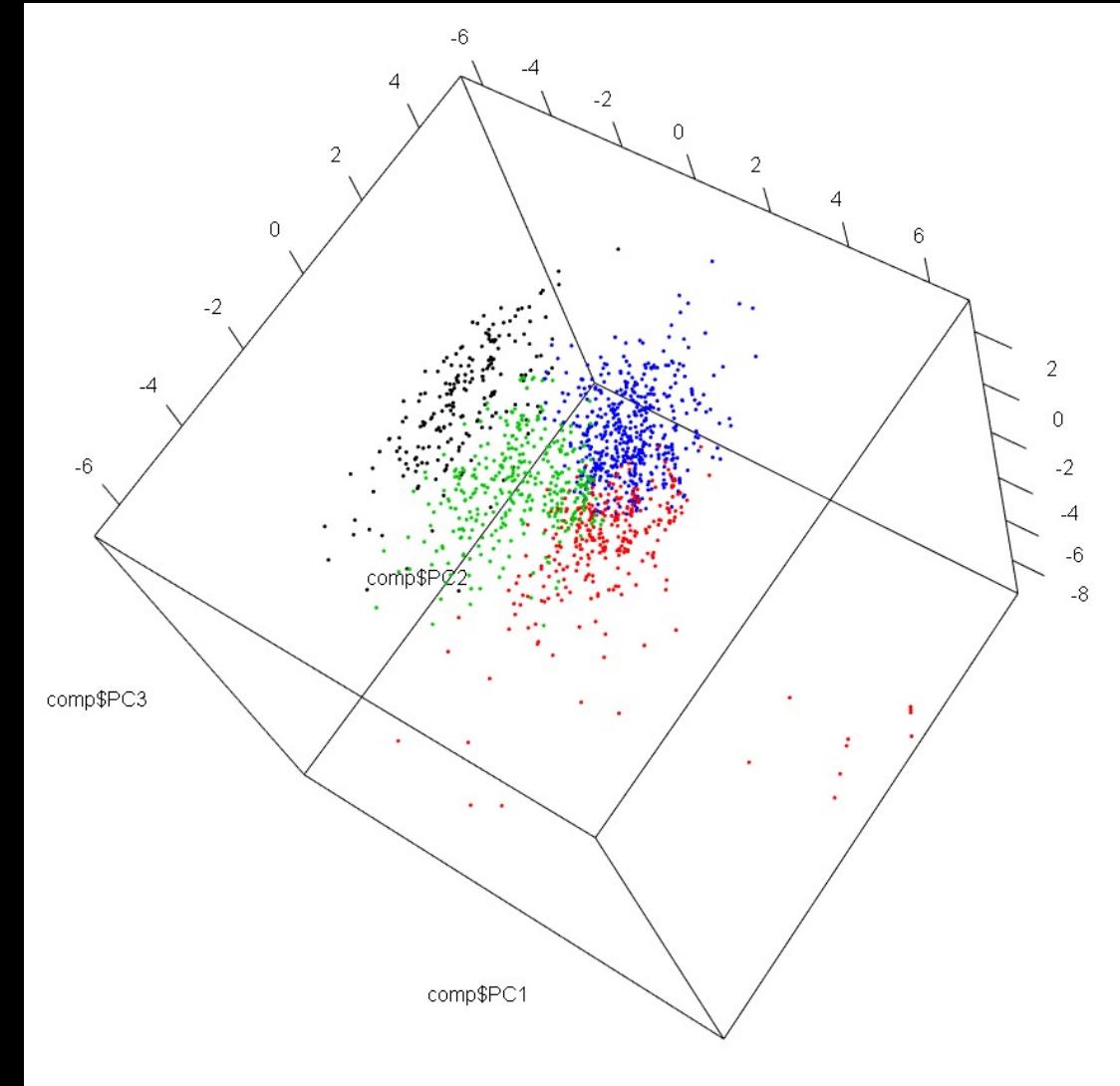


# Cluster by Position



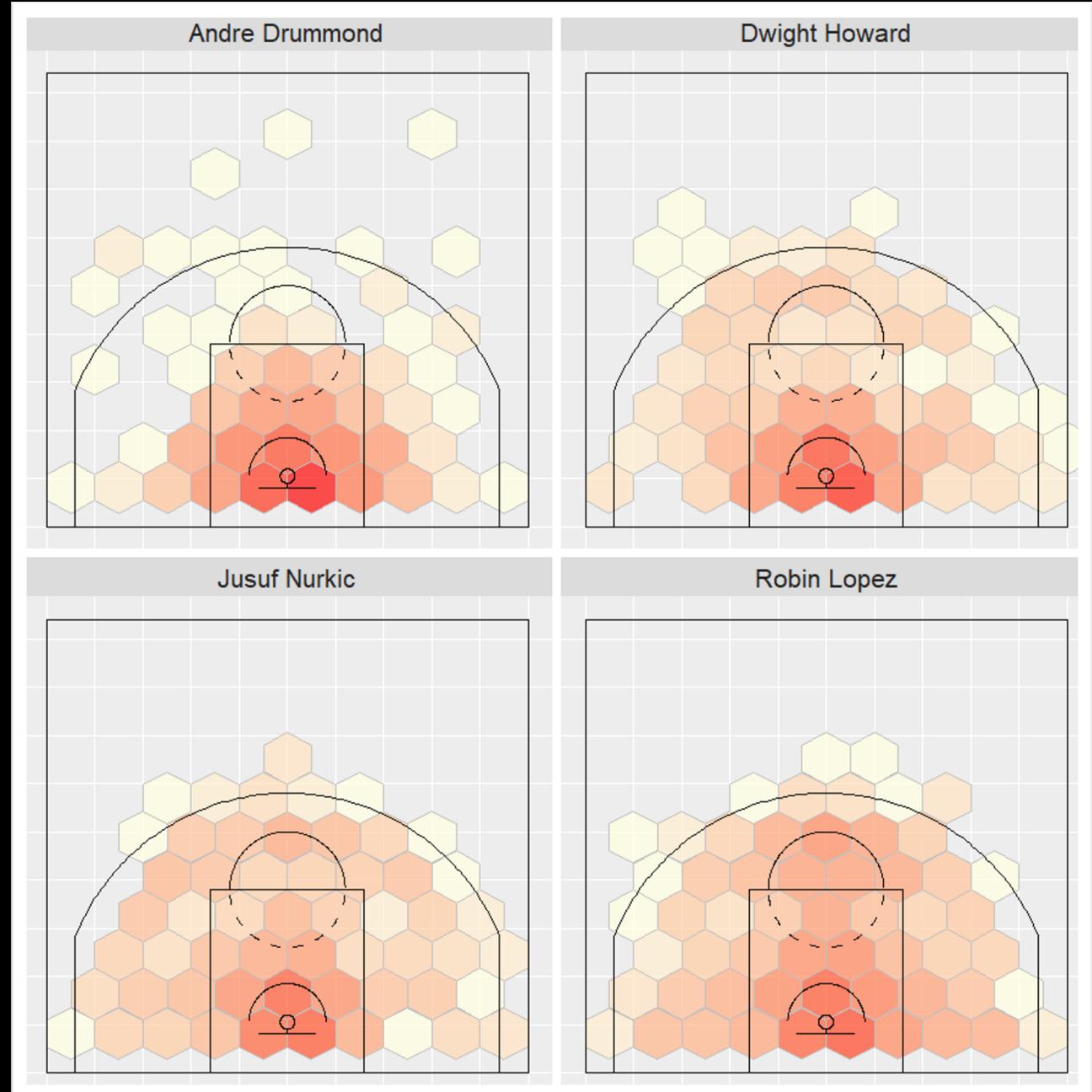


**HAC Clustering**  
**Euclidean Distancing**

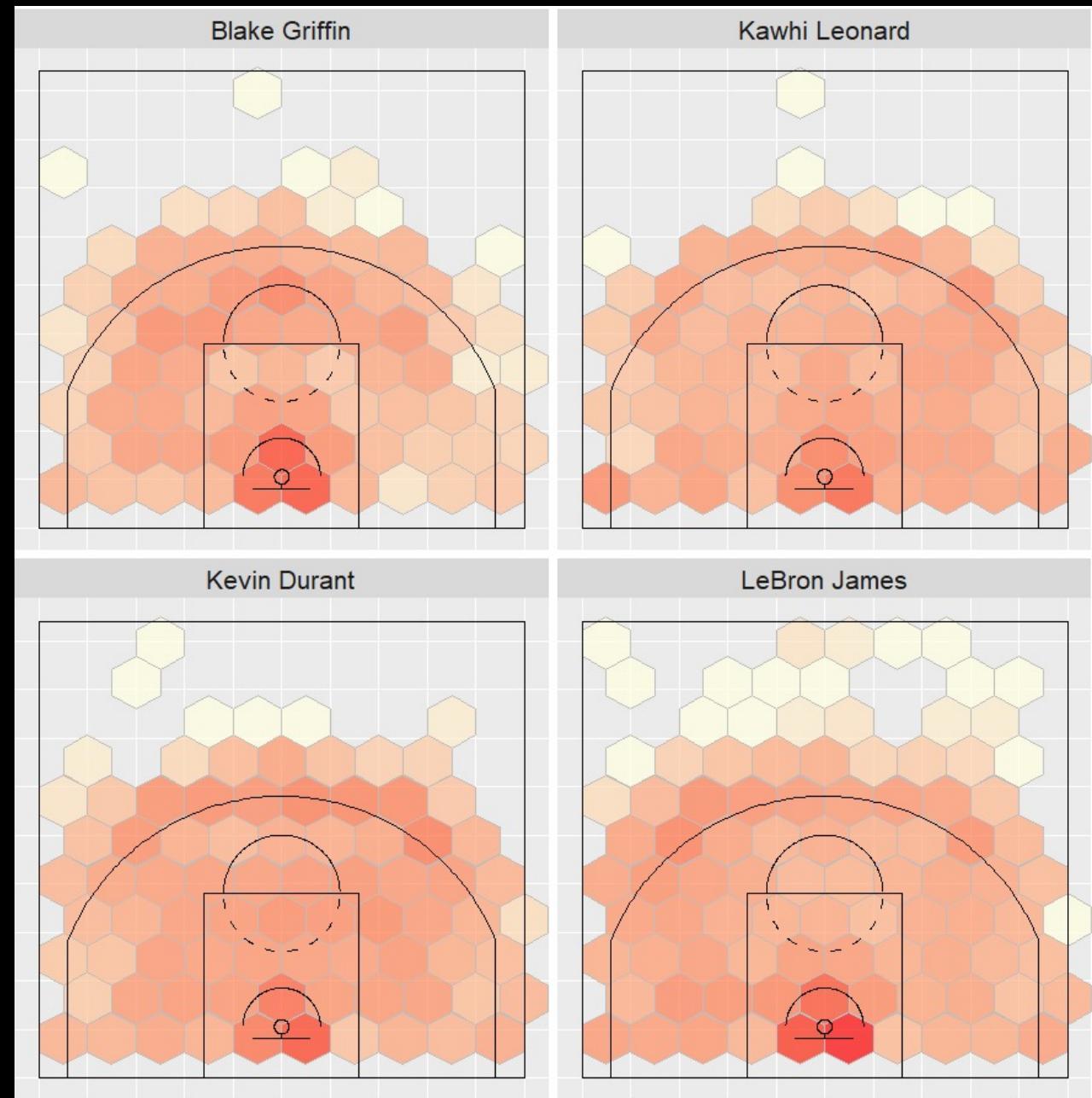


**K-Means Clustering**

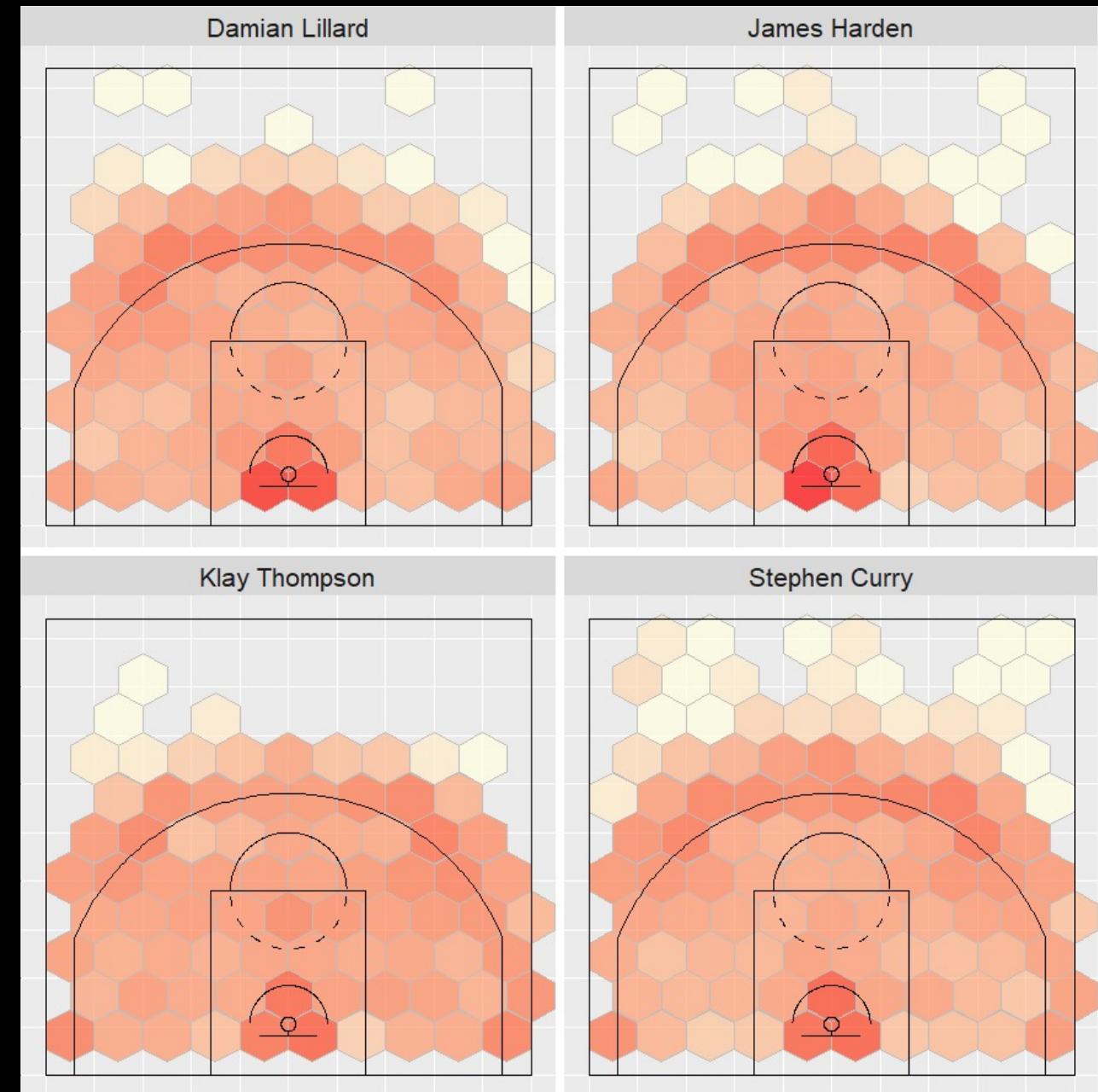
# Player Type One



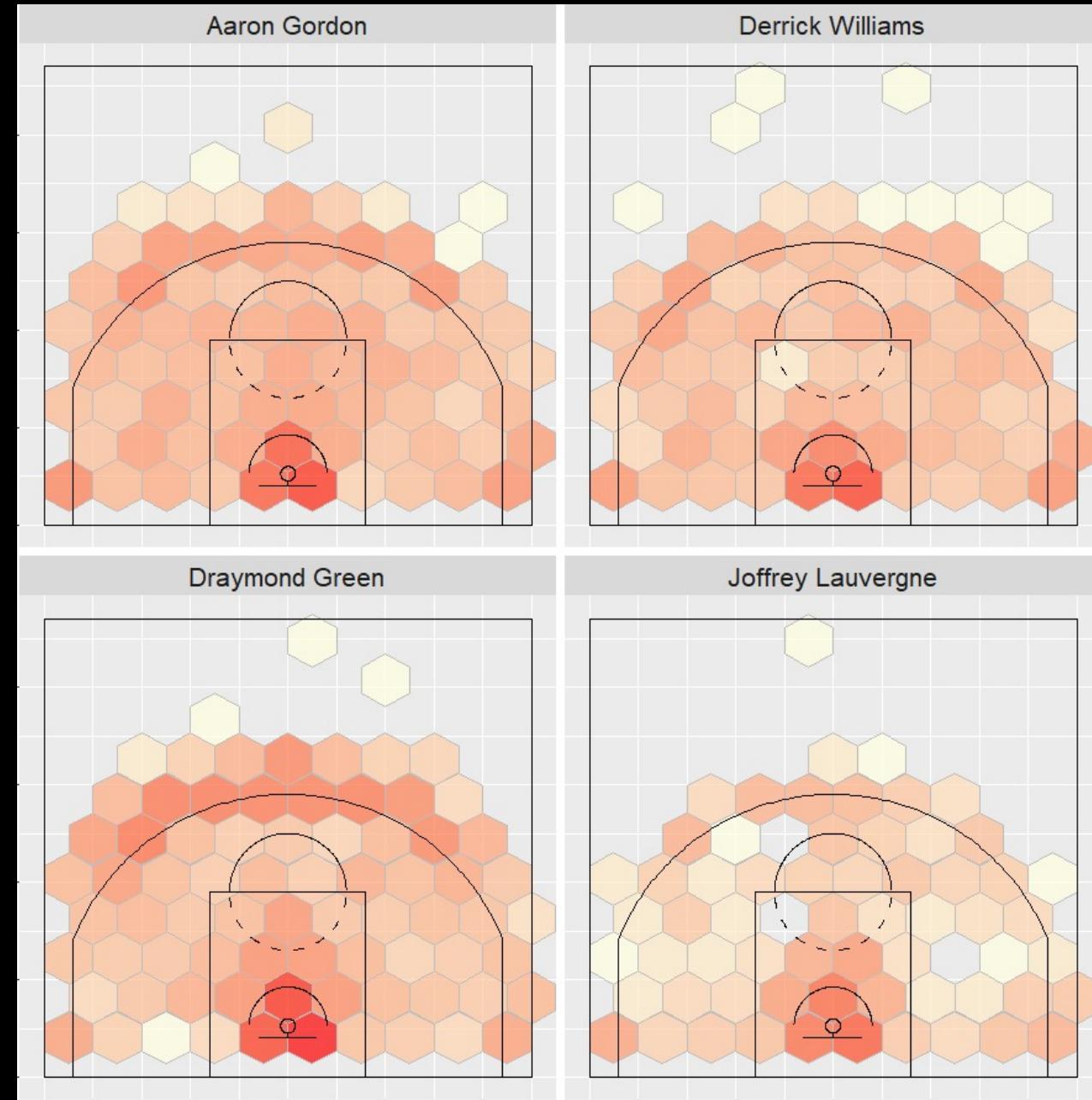
# Player Type Two



# Player Type Three



# Player Type Four

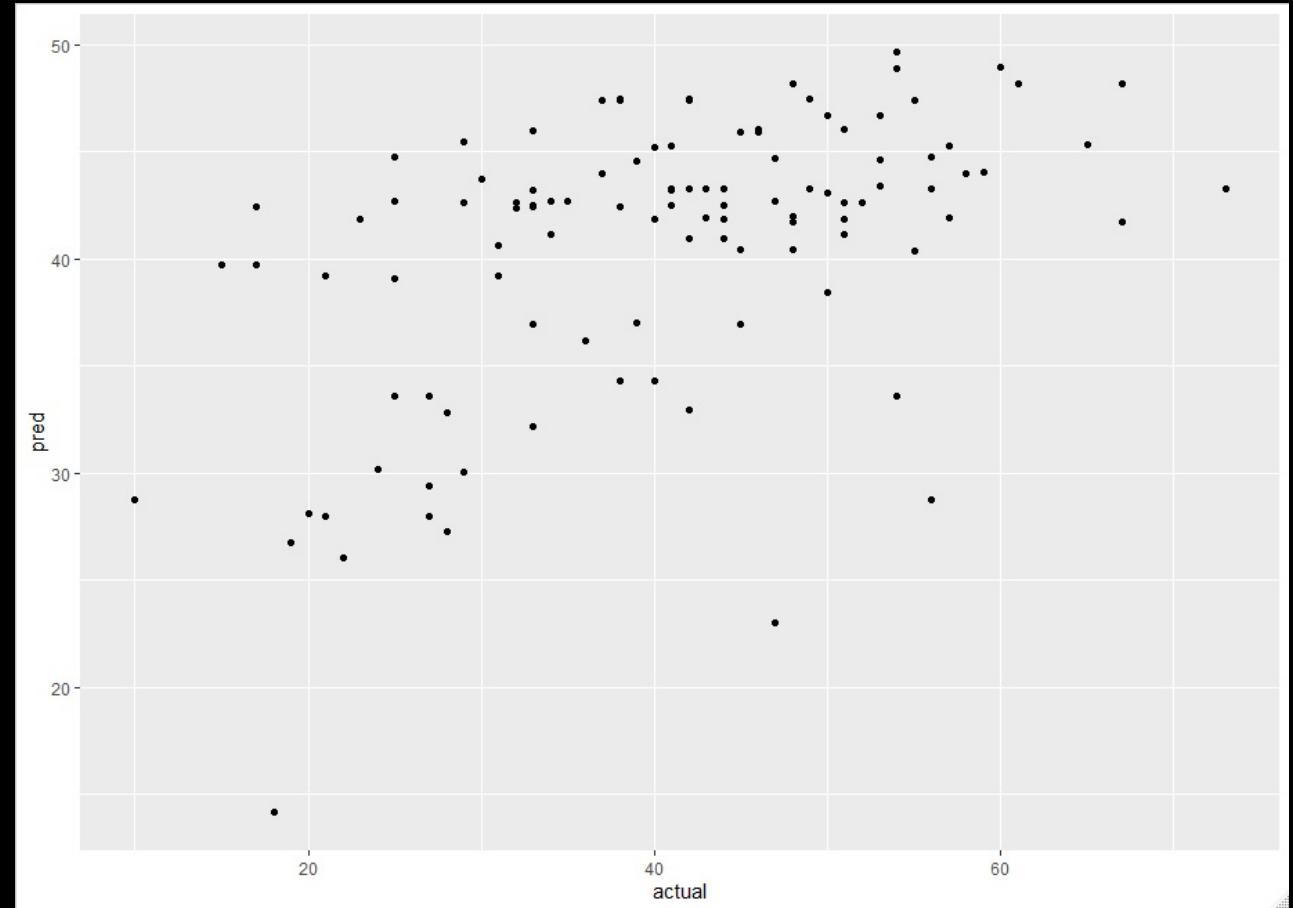


# Evaluation of Clustering Model

- Of 187 players with  $\geq 3$  seasons of data 84% were clustered as same type each year
- Number of each player type per roster correlated with season wins in linear model

## HAC Clustering

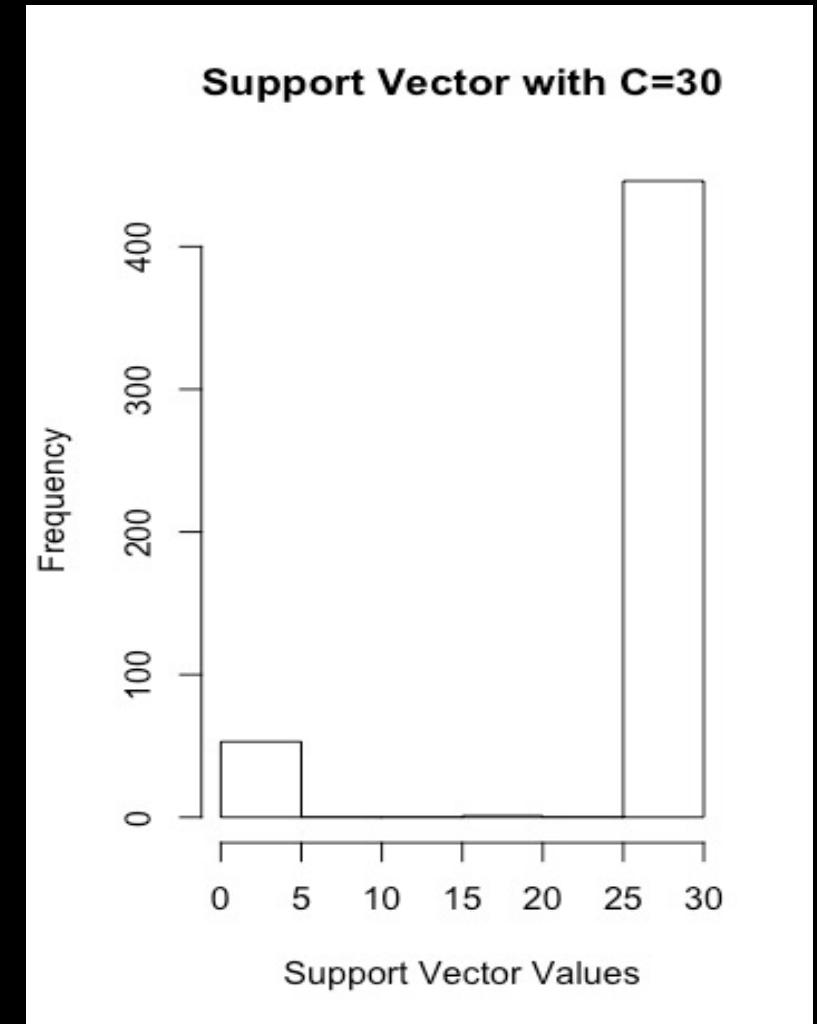
- RMSE: 11.3



# Support Vector Machine Analysis

Predicting a win or loss depending on home/away location status and opponent:

- Right-side shows difficulty in the model's predictive zoning
- Left-side shows points that are too easy to predict so offer no help to the modeling power
- Important to utilize the cost coefficient (C) to find a compromise that works for the model



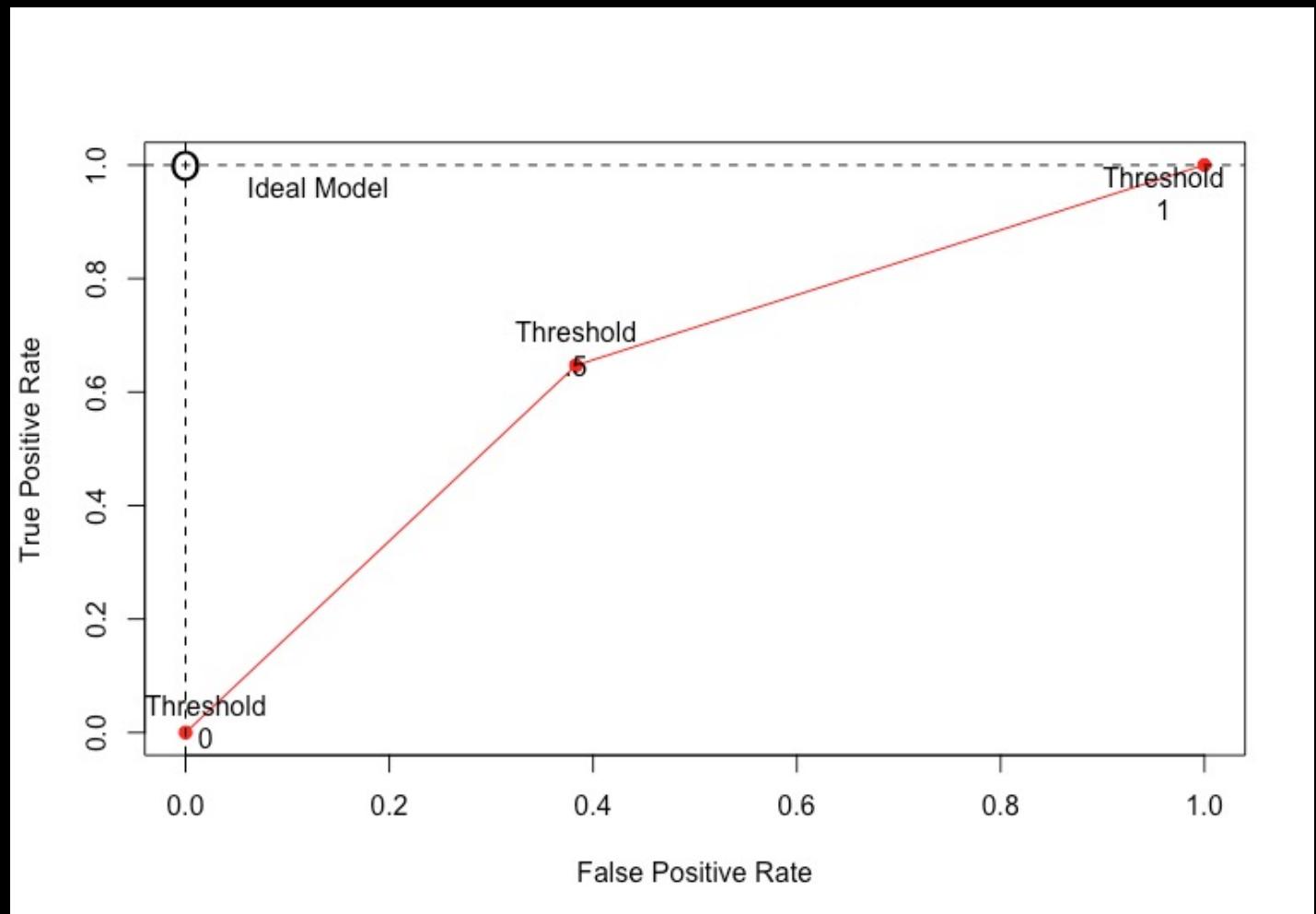
# Confusion Matrix and Statistics

<i>Prediction</i>	<i>L</i>	<i>W</i>
<i>L</i>	77	69
<i>W</i>	42	111

Accuracy: 62.9%

No Information Rate: 60.2%

Balanced Accuracy: 63.2%



# Support Vector Machine Analysis

Predicting a successful shot attempt depending on situational variables:

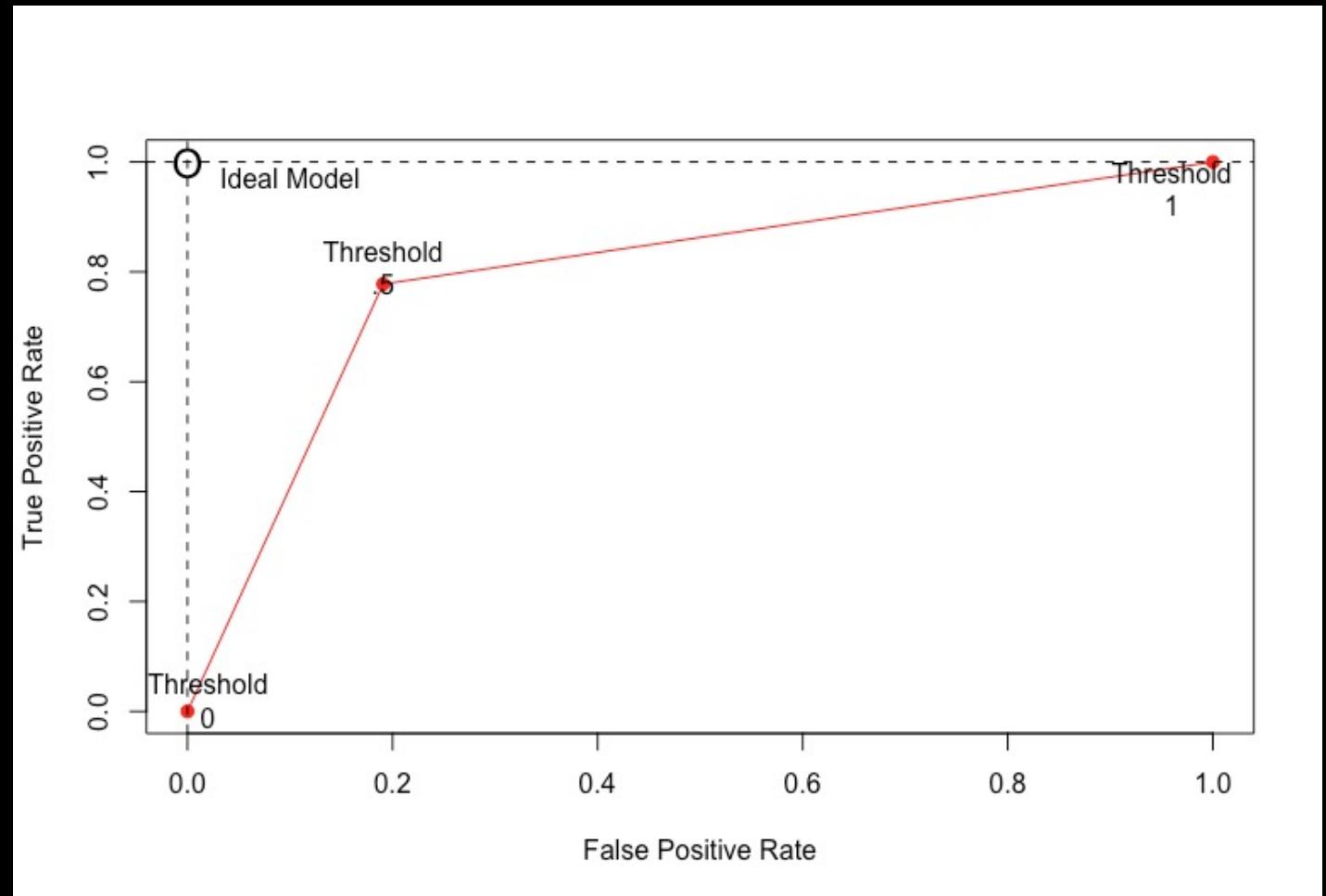
<i>Prediction</i>	<i>made</i>	<i>missed</i>
<i>made</i>	103	28
<i>missed</i>	34	134

Accuracy: 79.3%

No Information Rate: 54.2%

Balanced Accuracy: 79.0%

P-Value: 2e-16



# Decision Tree – Shot Result

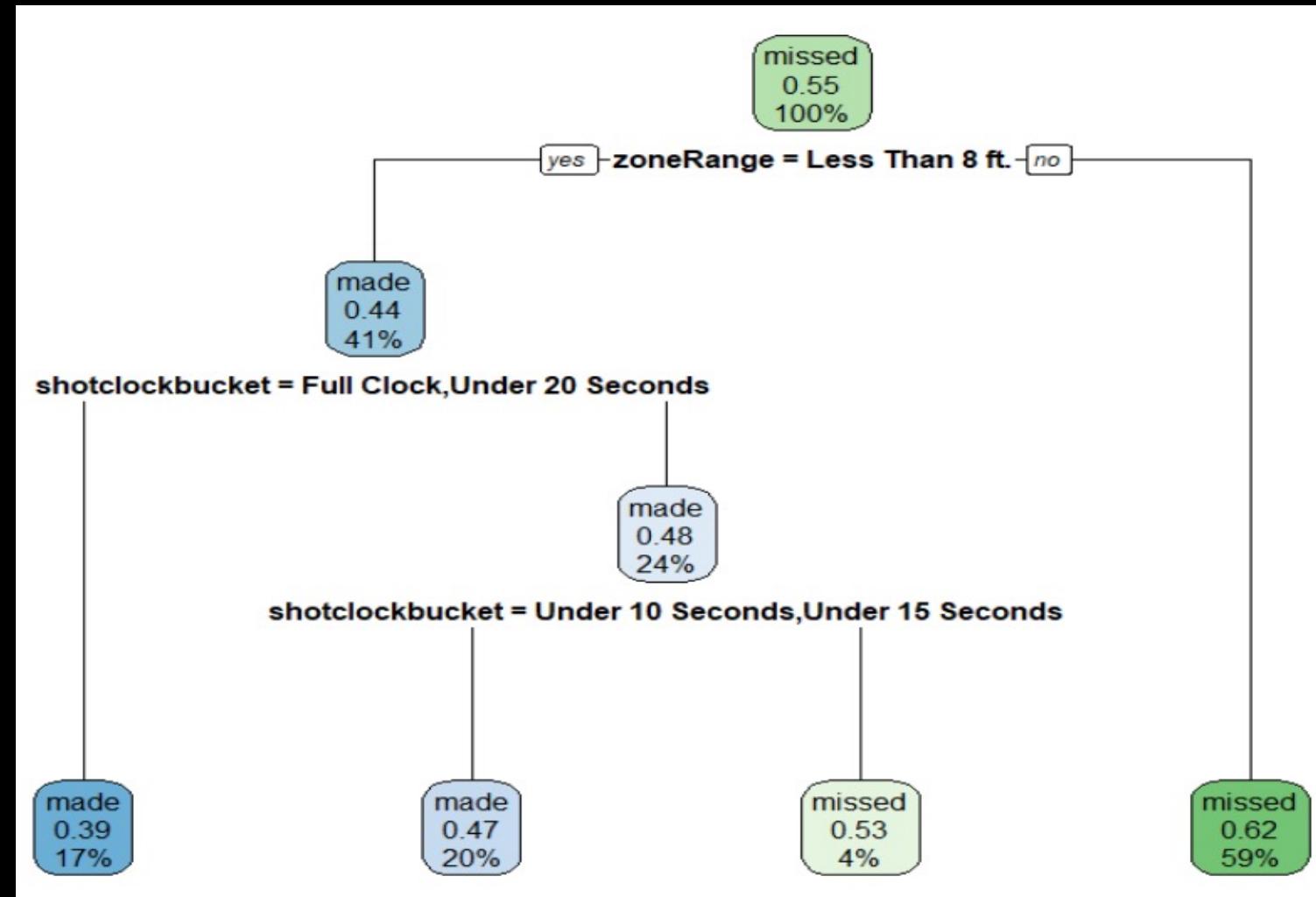
Variable Importance:

1. ZoneRange
2. SHOT\_DISTkbucket
3. shotclockbucket

## Confusion Matrix

	<i>made</i>	<i>missed</i>	<i>Sum</i>
<i>made</i>	<b>6,334</b>	4,876	11,210
<i>missed</i>	7,478	<b>11,838</b>	19,316
<i>Sum</i>	13,812	16,714	30,526

Prediction Accuracy: 59.53%



# Decision Tree - Shot Result

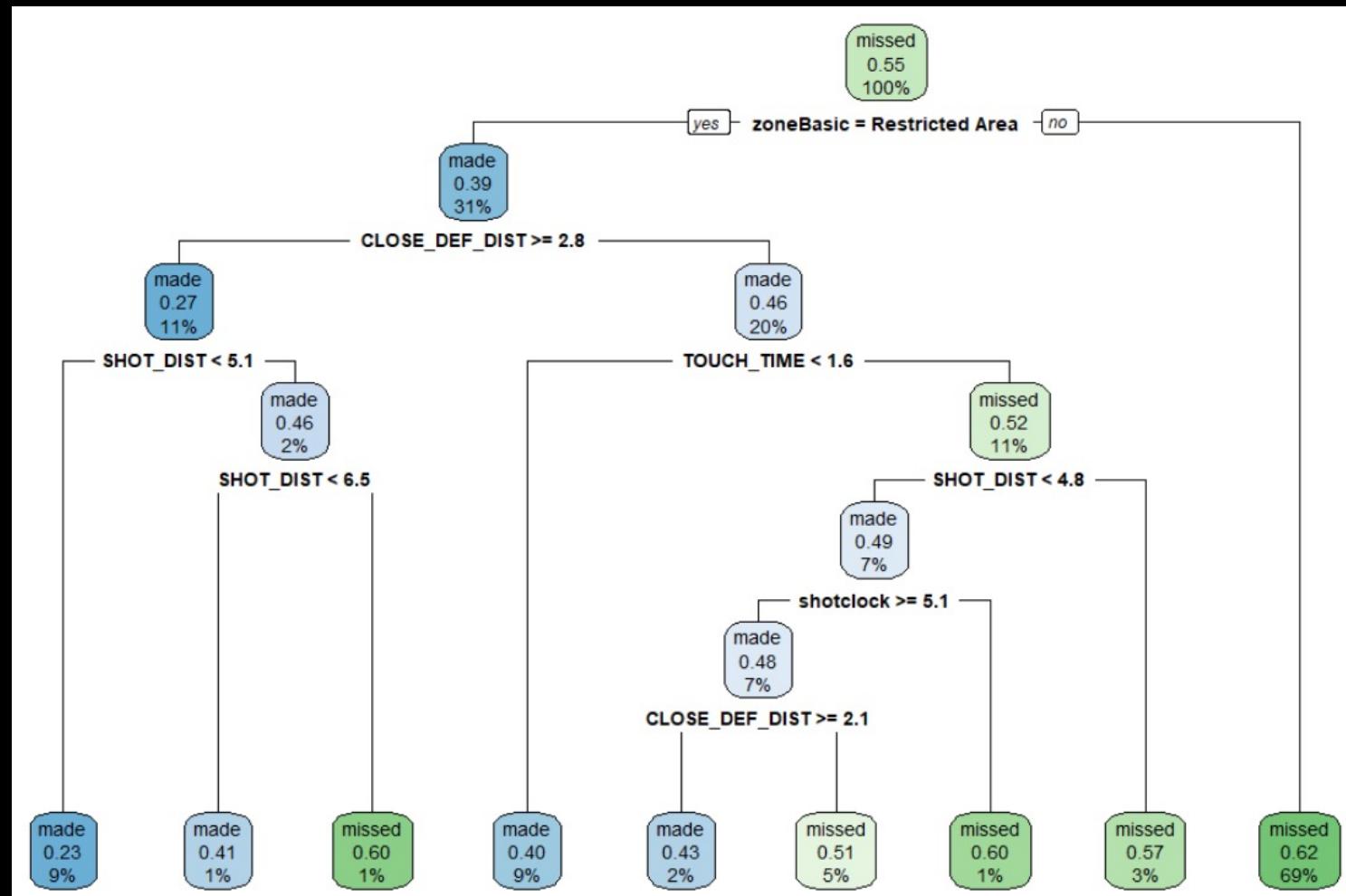
## Variable Importance:

1. ZoneBASIC
  2. SHOT\_DIST
  3. ZoneRange
  4. CLOSE\_DEF\_DIST
  5. Shotclock
  6. TOUCH\_TIME
  7. Dribbles

# Confusion Matrix

	<i>made</i>	<i>missed</i>	<i>Sum</i>
<i>made</i>	<b>4,535</b>	2,207	6,742
<i>missed</i>	9,277	<b>14,507</b>	23,784
<i>Sum</i>	13,812	16,714	30,526

Prediction Accuracy: 62.38%



# Decision Tree – Win/Loss

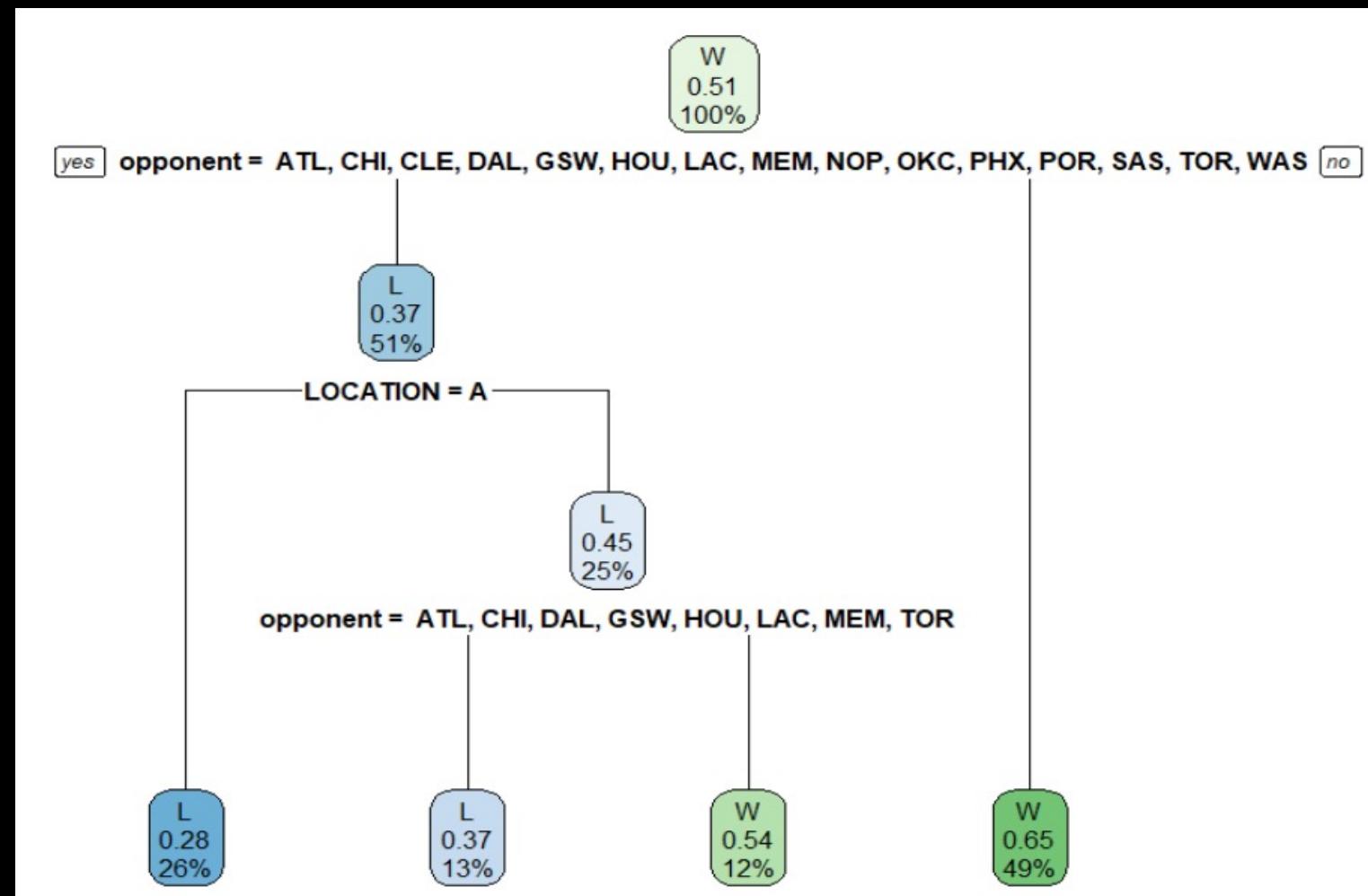
Variable Importance:

1. Opponent
2. Location

## Confusion Matrix

	<i>L</i>	<i>W</i>	<i>Sum</i>
<i>L</i>	<b>8,102</b>	3,734	11,836
<i>W</i>	7,093	<b>11,597</b>	18,690
<i>Sum</i>	15,195	15,331	30,526

Prediction Accuracy: 64.54%



# Association Rule Mining

- Unspecified "RHS"
- Irrelevant columns included
- Didn't give us any useful information about our data

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{SHOT_DIST=Mid Range, PTS_TYPE=3, CLOSE_DEF_DIST=Very Close, yearSeason=2015}	=> {zoneRange=24+ ft.}	0.2528480	0.9930828	0.2546092	3.785650	31872
[2]	{SHOT_DIST=Mid Range, PTS_TYPE=3, CLOSE_DEF_DIST=Very Close, slugSeason=2014-15}	=> {zoneRange=24+ ft.}	0.2528480	0.9930828	0.2546092	3.785650	31872
[3]	{SHOT_DIST=Mid Range, PTS_TYPE=3, CLOSE_DEF_DIST=Very Close, yearSeason=2015, slugSeason=2014-15}	=> {zoneRange=24+ ft.}	0.2528480	0.9930828	0.2546092	3.785650	31872
[4]	{SHOT_DIST=Mid Range, PTS_TYPE=3, yearSeason=2015}	=> {zoneRange=24+ ft.}	0.2580284	0.9930691	0.2598293	3.785597	32525
[5]	{SHOT_DIST=Mid Range, PTS_TYPE=3, slugSeason=2014-15}	=> {zoneRange=24+ ft.}	0.2580284	0.9930691	0.2598293	3.785597	32525
[6]	{SHOT_DIST=Mid Range, PTS_TYPE=3, yearSeason=2015, slugSeason=2014-15}	=> {zoneRange=24+ ft.}	0.2580284	0.9930691	0.2598293	3.785597	32525
[7]	{SHOT_DIST=Mid Range, PTS_TYPE=3, CLOSE_DEF_DIST=Very Close}	=> {zoneRange=24+ ft.}	0.2528480	0.9927426	0.2546965	3.784353	31872
[8]	{SHOT_DIST=Mid Range, PTS_TYPE=3}	=> {zoneRange=24+ ft.}	0.2580284	0.9927357	0.2599165	3.784326	32525
[9]	{SHOT_DIST=Mid Range, CLOSE_DEF_DIST=Very Close, zoneRange=24+ ft.}	=> {PTS_TYPE=3}	0.2528480	0.9985588	0.2532130	3.772300	31872
[10]	{SHOT_DIST=Mid Range, CLOSE_DEF_DIST=Very Close, yearSeason=2015, zoneRange=24+ ft.}	=> {PTS_TYPE=3}	0.2528480	0.9985588	0.2532130	3.772300	31872
[11]	{SHOT_DIST=Mid Range, CLOSE_DEF_DIST=Very Close, slugSeason=2014-15, zoneRange=24+ ft.}	=> {PTS_TYPE=3}	0.2528480	0.9985588	0.2532130	3.772300	31872
[12]	{SHOT_DIST=Mid Range, CLOSE_DEF_DIST=Very Close, yearSeason=2015, slugSeason=2014-15, zoneRange=24+ ft.}	=> {PTS_TYPE=3}	0.2528480	0.9985588	0.2532130	3.772300	31872
[13]	{SHOT_DIST=Mid Range, zoneRange=24+ ft.}	=> {PTS_TYPE=3}	0.2580284	0.9985570	0.2584013	3.772293	32525
[14]	{SHOT_DIST=Mid Range, yearSeason=2015, zoneRange=24+ ft.}	=> {PTS_TYPE=3}	0.2580284	0.9985570	0.2584013	3.772293	32525
[15]	{SHOT_DIST=Mid Range, slugSeason=2014-15,}	=> {PTS_TYPE=3}	0.2580284	0.9985570	0.2584013	3.772293	32525

# Association Rule Mining – Missed Shots (RHS)

Created a "test" dataset and removed columns:

- FGM
- PTS
- typeEvent
- isShotAttempted
- isShotMade

Specified "SHOT\_Result = missed" as the rhs of the apriori

lhs	rhs	support	confidence	coverage	lift	count
[1] {CLOSE_DEF_DIST=Very Close, typeAction=Jump Shot, nameZone=Back Court}	=> {SHOT_RESULT=missed}	0.00177	0.991	0.00178	1.81	223
[2] {CLOSE_DEF_DIST=Very Close, typeAction=Jump Shot, zoneRange=Back Court Shot}	=> {SHOT_RESULT=missed}	0.00177	0.991	0.00178	1.81	223
[3] {CLOSE_DEF_DIST=Very Close, typeAction=Jump Shot, zoneBasic=Backcourt}	=> {SHOT_RESULT=missed}	0.00153	0.990	0.00155	1.81	193
[4] {typeAction=Jump Shot, nameZone=Back Court}	=> {SHOT_RESULT=missed}	0.00184	0.987	0.00186	1.80	232
[5] {typeAction=Jump Shot, zoneRange=Back Court Shot}	=> {SHOT_RESULT=missed}	0.00184	0.987	0.00186	1.80	232
[6] {typeAction=Jump Shot, nameZone=Back Court, zoneRange=Back Court Shot}	=> {SHOT_RESULT=missed}	0.00184	0.987	0.00186	1.80	232
[7] {yearSeason=2015, typeAction=Jump Shot, nameZone=Back Court}	=> {SHOT_RESULT=missed}	0.00184	0.987	0.00186	1.80	232
[8] {slugSeason=2014-15, typeAction=Jump Shot, nameZone=Back Court}	=> {SHOT_RESULT=missed}	0.00184	0.987	0.00186	1.80	232
[9] {yearSeason=2015, typeAction=Jump Shot, zoneRange=Back Court Shot}	=> {SHOT_RESULT=missed}	0.00184	0.987	0.00186	1.80	232
[10] {slugSeason=2014-15, typeAction=Jump Shot, zoneRange=Back Court Shot}	=> {SHOT_RESULT=missed}	0.00184	0.987	0.00186	1.80	232
[11] {PTS_TYPE=3, typeAction=Jump Shot, nameZone=Back Court}	=> {SHOT_RESULT=missed}	0.00183	0.987	0.00186	1.80	231
[12] {PTS_TYPE=3, typeAction=Jump Shot, zoneRange=Back Court Shot}	=> {SHOT_RESULT=missed}	0.00183	0.987	0.00186	1.80	231
[13] {shotclock=Buzzer Beater, typeAction=Jump Shot, nameZone=Back Court}	=> {SHOT_RESULT=missed}	0.00178	0.987	0.00181	1.80	225
[14] {typeAction=Jump Shot, nameZone=Back Court, secondsRemaining=End Quarter}	=> {SHOT_RESULT=missed}	0.00178	0.987	0.00181	1.80	225
[15] {shotclock=Buzzer Beater, typeAction=Jump Shot, zoneRange=Back Court Shot}	=> {SHOT_RESULT=missed}	0.00178	0.987	0.00181	1.80	225
[16] {typeAction=Jump Shot, zoneRange=Back Court Shot, secondsRemaining=End Quarter}	=> {SHOT_RESULT=missed}	0.00178	0.987	0.00181	1.80	225
[17] {typeAction=Jump Shot, minutesRemaining=0, nameZone=Back Court}	=> {SHOT_RESULT=missed}	0.00178	0.987	0.00180	1.80	224
[18] {typeAction=Jump Shot, nameZone=Back Court}	=> {SHOT_RESULT=missed}	0.00178	0.987	0.00180	1.80	224

# Association Rule Mining – Made Shots (RHS)

Created a "test" dataset and removed columns:

- FGM
- PTS
- typeEvent
- isShotAttempted
- isShotMade

Specified "SHOT\_Result = made" as the rhs of the apriori

[1]	{TOUCH_TIME=0.2, typeAction=Alley Oop Dunk Shot} => {SHOT_RESULT=made} 0.00101	1.000	0.00101	2.21	127
[2]	{TOUCH_TIME=0.2, typeAction=Alley Oop Dunk Shot, zoneBasic=Restricted Area} => {SHOT_RESULT=made} 0.00101	1.000	0.00101	2.21	127
[3]	{TOUCH_TIME=0.2, typeAction=Alley Oop Dunk Shot, zoneRange=Less Than 8 ft.} => {SHOT_RESULT=made} 0.00101	1.000	0.00101	2.21	127
[4]	{TOUCH_TIME=0.2, SHOT_DIST=Short Range, typeAction=Alley Oop Dunk Shot} => {SHOT_RESULT=made} 0.00101	1.000	0.00101	2.21	127
[5]	{TOUCH_TIME=0.2, typeAction=Alley Oop Dunk Shot, nameZone=Center} => {SHOT_RESULT=made} 0.00101	1.000	0.00101	2.21	127
[6]	{TOUCH_TIME=0.2, PTS_TYPE=2, typeAction=Alley Oop Dunk Shot} => {SHOT_RESULT=made} 0.00101	1.000	0.00101	2.21	127
[7]	{TOUCH_TIME=0.2, CLOSE_DEF_DIST=Very Close, typeAction=Alley Oop Dunk Shot} => {SHOT_RESULT=made} 0.00101	1.000	0.00101	2.21	127
[8]	{TOUCH_TIME=0.2, yearSeason=2015, typeAction=Alley Oop Dunk Shot} => {SHOT_RESULT=made} 0.00101	1.000	0.00101	2.21	127
[9]	{TOUCH_TIME=0.2, slugSeason=2014-15, typeAction=Alley Oop Dunk Shot} => {SHOT_RESULT=made} 0.00101	1.000	0.00101	2.21	127
[10]	{PTS_TYPE=2, typeAction=Pullup Jump shot, slugTeamHome=WAS} => {SHOT_RESULT=made} 0.00102	0.992	0.00103	2.19	129
[11]	{typeAction=Slam Dunk Shot, secondsRemaining=End Quarter} => {SHOT_RESULT=made} 0.00149	0.989	0.00151	2.19	188
[12]	{typeAction=Slam Dunk Shot, zoneBasic=Restricted Area, secondsRemaining=End Quarter} => {SHOT_RESULT=made} 0.00149	0.989	0.00151	2.19	188
[13]	{typeAction=Slam Dunk Shot, zoneRange=Less Than 8 ft., secondsRemaining=End Quarter} => {SHOT_RESULT=made} 0.00149	0.989	0.00151	2.19	188
[14]	{SHOT_DIST=Short Range, typeAction=Slam Dunk Shot, secondsRemaining=End Quarter} => {SHOT_RESULT=made} 0.00149	0.989	0.00151	2.19	188
[15]	{typeAction=Slam Dunk Shot, nameZone=Center, secondsRemaining=End Quarter} => {SHOT_RESULT=made} 0.00149	0.989	0.00151	2.19	188
[16]	{PTS_TYPE=2, typeAction=Slam Dunk Shot, secondsRemaining=End Quarter} => {SHOT_RESULT=made} 0.00149	0.989	0.00151	2.19	188
[17]	{yearSeason=2015, typeAction=Slam Dunk Shot, secondsRemaining=End Quarter} => {SHOT_RESULT=made} 0.00149	0.989	0.00151	2.19	188
[18]	{slugSeason=2014-15, typeAction=Slam Dunk Shot, secondsRemaining=End Quarter} => {SHOT_RESULT=made} 0.00149	0.989	0.00151	2.19	188

# Conclusions

- We found that the SVM analysis was the most effective in predicting whether a shot will be made or not.
- Decision Trees were the most effective in predicting a win vs. a loss, although the SVM was just slightly lower in predicting accuracy.
- There may be other attributes that could be valuable to this analysis, such as, a team's record during a particular game, injuries that could affect a team's game plan, travel time for the away team, etc.