California State University, Long Beach

The Hidden Secrets Behind Making the MLB Playoffs

STAT 550: Multivariate Analysis Final

Denisse Valdez
Ho Joon Lee
Kevin Merritt
5-5-2016

INTRODUCTION

According to USA Today Sports, Major League Baseball teams who make the playoffs earn on average $20 to $30 million more in revenue each year than teams who do not make the playoffs (Barmasse, 2014). Knowing this, the objective for this project was to determine which team performance statistics play the largest role in reaching the playoffs and thus increasing annual revenue. Identifying these factors is important since an extra 20 million dollars can improve a team's performance considerably either by affording new players or new management. Additionally, it is commonly known pitchers are the highest paid players in baseball. Therefore, do pitching statistics also play the largest role in determining if a team will make the playoffs?

DATA

This project uses MLB statistics from the years 2011 to 2015, which was collected from the Elias Sports Bureau (ESPN, 2016). The five most recently completed seasons were chosen because it was determined one season wasn't enough and consolidating the data for more than five seasons would be beyond the scope of this project. Consequently, each of the 30 MLB teams have five years of observations, resulting in a total of 150 data observations. From the wide selection of baseball statistics, we focused on 25 of the "major" statistics for batting, pitching, and fielding. PitchERA, BatAB/HR, PitchR, PitchBAA, and PitchBB are all variables which were considered "good" if they have lower values (highlighted in the table below). In addition, our final data set for this project included two categorical variables – team league and playoffs teams.

Therefore, a logistic analysis model is appropriate for our data set, where 0 means National League (NL) and 1 means American League (AL) for the variable league. For the variable playoffs, 1 means the team made the playoffs and 0 means they did not make the playoffs. Since we have 27 variables, the following is the list of the variables which require more in depth knowledge of baseball. The rest of the variable descriptions can be found in Appendix A1.

BatTB	total bases; H+ (# of 2B)*2 + (# of 3B)*3+ (# of HR)*4
BatRBI	runs batted in; A run scored because of the action of a batter
BatAVG	batting average; Hits divided by At Bats
BatOBP	on base percentage; (H + BB + HBP) divided by (AB + BB + HBP + SF) (Hits + walks + hit by pitch)/(at-bats + walks + hit by pitch + sacrifice flies)
BatSLG	slugging percentage; Total Bases divided by At Bats
BatOPS	OPS = OBP +SLG
BatAB/HR	at bats to home run ratio
PitchERA	Earned-run average; (ER times 9) divided by IP (Earned runs*9)/(innings pitched)
PitchSV	saves; Earned when a pitcher finishes a game without having given up the lead after entering in a save situation
PitchCG	complete games; the act of a pitcher pitching an entire game without the benefit

	of a relief pitcher.
PitchSHO	shutouts, refers to the act by which a single pitcher pitches a complete game and does not allow the opposing team to score a run.
PitchQS	quality starts; pitcher completes at least six innings and permits no more than three earned runs.
PitchBAA	opponents' batting average; measures a pitcher's ability to prevent hits during official at bats.
PitchSV%	save percentage; # of saves divided by save opportunities
PitchK/BB	strikeout to walk ratio
FieldFPCT	fielding percentage (PO + A) divided by (PO + A + E) (Putouts + Assists)/(Putouts + Assists + Errors)

Table 1: Variable Explanations Where Highlighted Variables are "Good" When Low Values

PRELIMINARY ANALYSIS

For this analysis, a correlation matrix was used because there is an inconsistency in the units used to measure each variable. Additionally, looking at the sample of simple statistics below (Table 2) for a few of the variables in the data set, it is clear the range of standard deviations is very large which confirms the choice of using the correlation matrix for analysis.

Simple Statistics									
	BatAB	BatR	BatH	Bat2B	Bat3B	BatHR	BatTB	BatRBI	BatAVG
Mean	5520.853333	683.2533333	1400.833333	275.0733333	29.23333333	154.9466667	2199.213333	649.7933333	0.2536400000
StD	73.578556	67.1739390	76.246370	24.3443134	8.86014882	31.5426177	150.088457	66.3135016	0.0112728694

Table 2: Sample of Basic Statistics for Variables in Data Set

In Appendix A2, the full correlation matrix for our data set can be found. In general, we found there to be a lot of variables highly correlated with each other. This is due to the fact certain variables play a role in others. For example, batter hits and batter average would directly affect the on base percentage.

To verify the assumption of normality, a QQ Plot (Figure 1) of the residuals was produced. Clearly the residuals generally form a straight line confirming the assumption of normality is not violated. Also, the residuals for each variable are randomly scattered about the zero line (Figure 2) indicating the assumption of equal variances is not violated. The rest of the residual plots can be seen in Appendix A3.

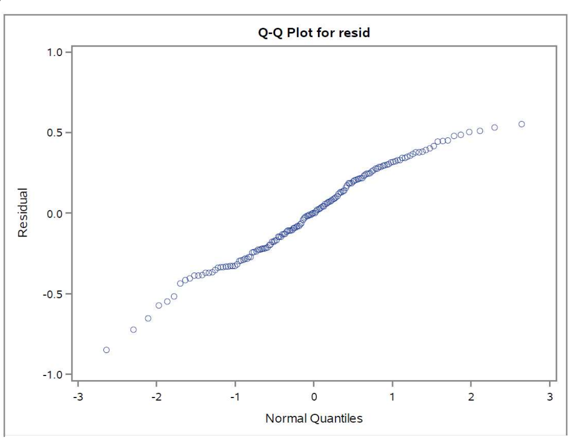


Figure 1: Q-Q Plot for Residuals

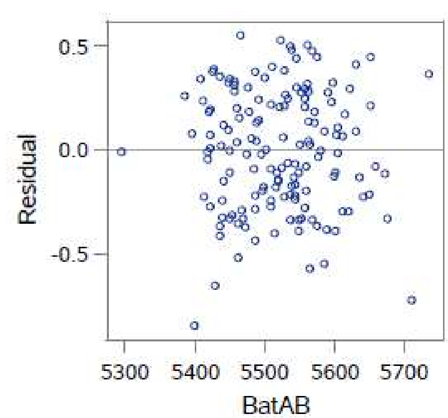


Figure 2: Residual vs BatAB Plot

PRINCIPAL COMPONENT ANALYSIS (PCA)

Due to the large number of variables in the data set, variable reduction is an essential part of the analysis. Based on the scree plot below (Figure 3), there is a significant drop in eigenvalues from one to two and two to three but the drop between three and four is not as significant. This is confirmed by the proportion and cumulative columns (Table 3). In general, a cumulative percentage of over 70% is considered okay; we decided 66% is appropriate for our data set due to the large number of observations.

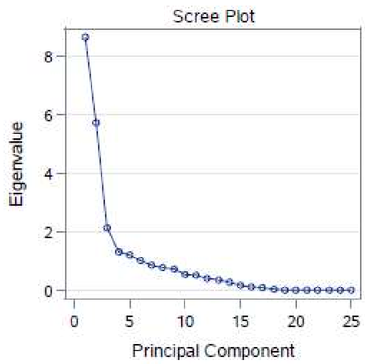


Figure 3: Scree Plot

Eigenvalues of the Correlation Matrix				
	Eigenvalue	Difference	Proportion	Cumulative
1	8.67369187	2.93581635	0.3469	0.3469
2	5.73787551	3.58364813	0.2295	0.5765
3	2.15422738	0.83521987	0.0862	0.6626
4	1.31900752	0.10029265	0.0528	0.7154
5	1.21871487	0.19590432	0.0487	0.7641

Table 3: Eigenvalue Table for First Five PCA's

The table below (Table 4) can be used to get a more accurate picture of what each principal component represents from the original variables. The following is an explanation of each principal component used:

Prin1: Measurement of hitting statistics for a particular team

Prin2: Measurement of pitching statistics for a particular team

Prin3: Measurement of "power ball[1]"

Keep in mind, some of the variables are "good" when they have large negative value within the principal component (circled in Figure 4).

	Prin1	Prin2	Prin3
BatAB	0.223957	-.086212	0.206371
BatR	0.321388	0.043724	-.026060
BatH	0.293195	-.048041	0.282776
Bat2B	0.234161	-.043604	0.235493
Bat3B	0.036711	-.091607	0.334377
BatHR	0.224519	0.075929	-.478050
BatTB	0.332816	0.005578	-.080073
BatRBI	0.322248	0.050583	-.034990
BatAVG	0.289824	-.032268	0.283321
BatOBP	0.288449	0.064135	0.199139
BatSLG	0.325247	0.026639	-.138105
BatOPS	0.333172	0.040955	-.031302

	Prin1	Prin2	Prin3
BatAB/HR	-.206920	-.082150	0.487704
PitchERA	0.058615	-.377700	-.115868
PitchSV	-.003077	0.269394	0.042271
PitchCG	0.018252	0.132086	0.071276
PitchSHO	-.021406	0.275158	0.169993
PitchQS	0.038196	-.020946	-.144585
PitchR	0.053906	-.382261	-.100183
PitchBB	-.022349	-.282650	-.003093
PitchSO	0.002557	0.278387	-.012702
PitchBAA	0.035518	-.354040	-.014738
PitchSV%	-.031129	0.264306	0.072194
PitchK/BB	0.016329	0.346530	-.013431
FieldFPCT	0.031338	0.166682	-.097924

Table 4: Loadings for Principal Components 1, 2, and 3

[1] Power ball refers to scoring runs based off of extra base hits such as 3B and HR

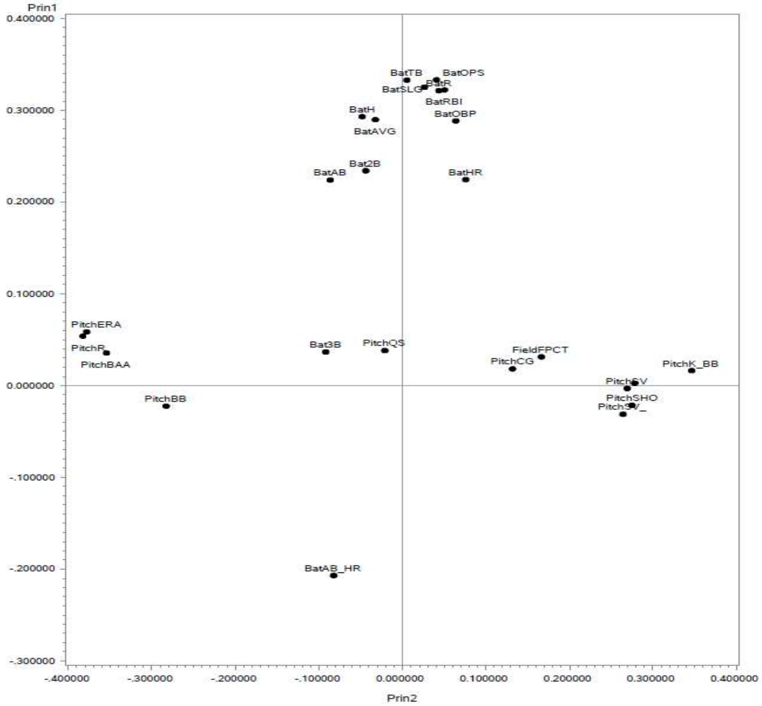


Figure 4: Prin1 vs. Prin2 with "Good" Variables Circled

Our next step in the analysis was to compare the graphs of prin1 vs prin2, prin1 vs prin3, and prin2 vs prin3. In doing so, teams were plotted in each combination by league where red is the AL and blue is the NL. The principal component plots by league shows no particular league is better in any particular principal component. For example, both leagues are randomly scattered when it comes to hitting (prin1), pitching (prin2), and "power ball" (prin3). These plots can be seen in Appendix A4.

Next, the same combinations of principal components were plotted by whether a team reached the playoffs (red) or not (blue). It is clear pitching (prin2) is the best predictor for whether a team will make the playoffs or not. This can be seen in Figure 6. Hitting (prin1) also plays a role in predicting if a team makes the playoffs (Figure 5). "Power ball" (prin3) doesn't seem to have an effect on playoffs when coupled with prin1 either (figure in Appendix A4).

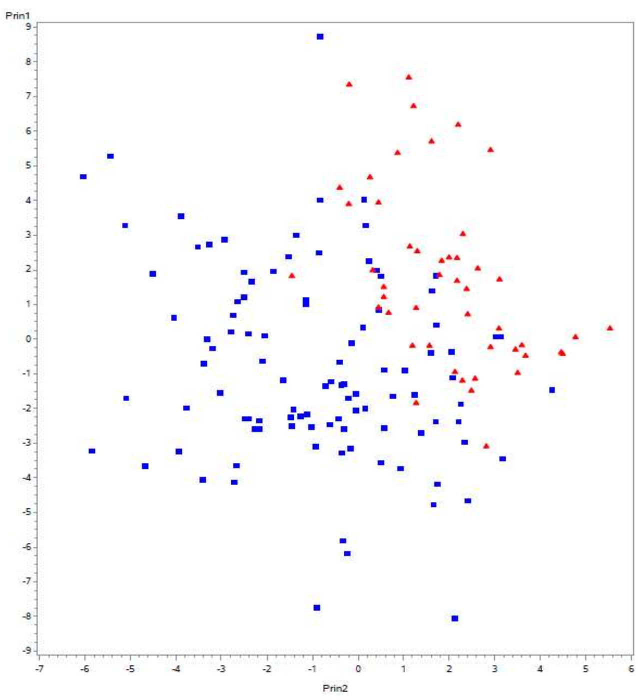


Figure 5: Prin1 vs Prin2 Separated by Playoffs

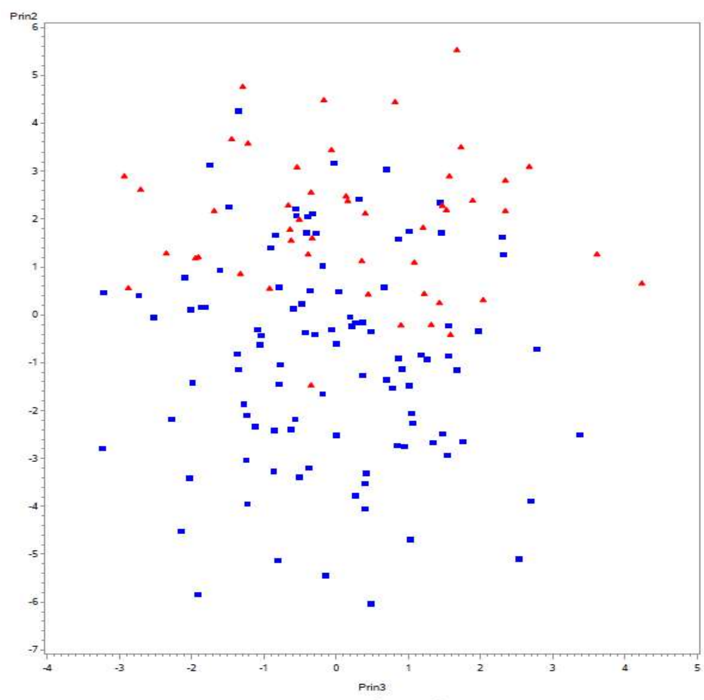


Figure 6: Prin2 vs Prin3 Separated by Playoffs

CLUSTER ANALYSIS (CA)

Based on the PC analysis performed above, we saw prin1 and prin2 are the most important components in determining if a team will make the playoffs. Therefore, our group performed further analysis of these two components using cluster analysis. The number of clusters used in the analysis was determined based on the tree diagram located in Appendix A5. Based on this diagram, we chose five clusters to group the variables.

Cluster analysis classifies the teams based on their statistics to groups which are similar. The first cluster plot (Figure 7) below, shows all the teams based on prin1 and prin2 and whether they made the playoffs or not. The green and red clusters in the plot show teams which made the playoffs. The black, orange, and blue indicate teams who did not make the playoffs. This coincides with the results found in our PCA section.

The second plot (Figure 8) shows prin1 and prin2 with the team names as their indicator. Knowing red and green indicate teams who made the playoffs and black, orange, and blue are teams who did not make the playoffs, we are able to make general assumptions about teams. For example, Colorado (all five years in the black cluster) tends to focus on hitting instead of pitching and it has resulted in not making the playoffs in the last five years. Another good illustration of the fine line between making the playoffs and not is seen in Atlanta11 and Atlanta12. In both seasons, Atlanta had similar pitching statistics but a slight increase in their hitting performance in 2012 caused them to make the playoffs.

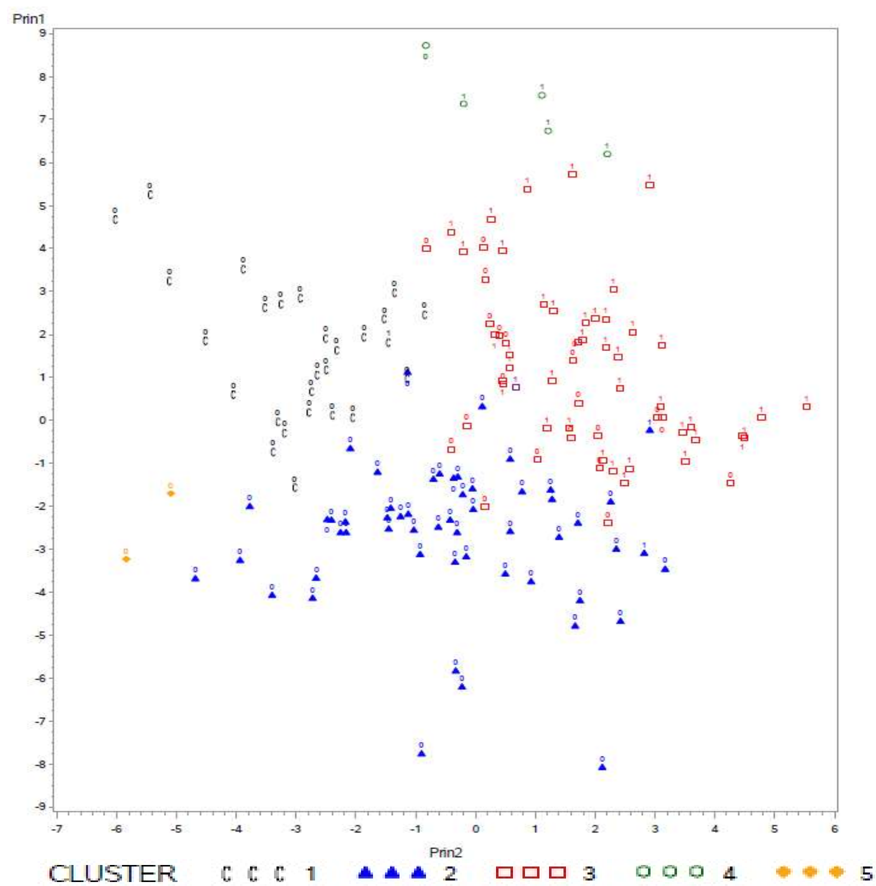


Figure 7: Prin1 vs. Prin2 CA Labeled with Playoffs

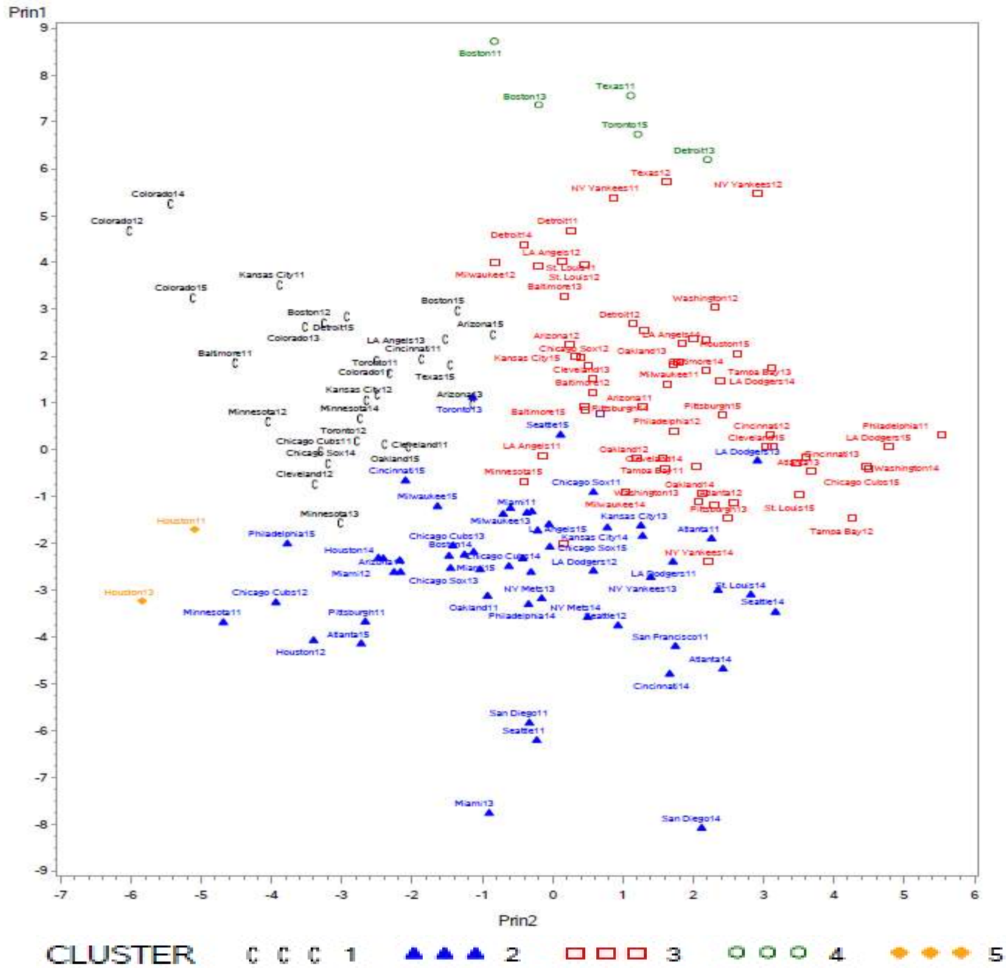


Figure 8: Prin1 vs. Prin2 CA with Team Name Labels

LOGISTIC ANALYSIS

Since we have analyzed the background information of the data, our next logical step was to attempt to predict what teams will make the playoffs and which ones will go home early. We chose to use Logistical Analysis instead of discriminate analysis because the response variable in our case is categorical (binary). Since PCA was already used to reduce the number of dimensions in our data, the logistic analysis was run using the three components identified previously.

Looking at the Maximum Likelihood Estimates shows a p-value for prin3 as .0507 (Table 5). While this typically would indicate removal of the component at the alpha level of .05, we included it because during our PC analysis, we saw prin3 accounts for about 10% of the explanation of the variance in our model. The estimates show prin2 has a significant amount of impact on the model compared to prin1 and prin3.

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-2.1042	0.4445	22.4096	<.0001
Prin1	1	0.6502	0.1399	21.5986	<.0001
Prin2	1	1.2020	0.2273	27.9704	<.0001
Prin3	1	0.3649	0.1868	3.8178	0.0507

Table 5: Maximum Likelihood Estimates for Prin 1, 2, and 3

Cross validation was then used to check the adequacy of the model. Typically a decision rule of 50% is used when classifying but we chose to use 60% in this case. We chose this rule because it is riskier for a team to believe they will make the playoffs when they end up missing them. For example, a team which thinks they will make the playoffs will believe they have an extra 20 to 30 million dollars in revenue when they don't. Since only 1/3 of MLB teams make the playoffs every year, we decided to use a decision rule of 60%.Therefore, we felt it would be better to be conservative and predict a team will miss the playoffs when they actually make it.

Using the decision rule of 60%, there were a total of 21 misclassifications out of 150 which equates to a 14% misclassification rate (Table 6). Out of the 21, 14 of the misclassifications were teams predicted to miss the playoffs but actually made them (false-negative). This boils down to seven teams believing they would get this increase in revenue and not actually getting it (false-positive). We expected to see more false-negatives than false-positives which turned out to be true.

Table of newplayoffs by Playoffs			
	Playoffs(Playoffs)		
newplayoffs	0	1	Total
0	95 63.33 87.16 93.14	14 9.33 12.84 29.17	109 72.67
1	7 4.67 17.07 6.86	34 22.67 82.93 70.83	41 27.33
Total	102 68.00	48 32.00	150 100.00

Table 6: Results of Logistic Analysis

In more detail, out of the 48 teams which made the playoffs, 14 of them were originally projected to miss them. This equates to 29.17% of teams who ended up making the playoffs thinking they wouldn't. On the other hand, out of the 102 teams which did not make the playoffs, seven of them believed they would but ended up missing them. This equates to 6.86% of those teams. Keep in mind our model was designed to favor misclassifying false-negative situations. For this reason, we believe our model is adequate and a pretty good representation of which teams will make the playoffs and which teams will end up missing out on the extra 20 to 30 million dollars in revenue.

CONCLUSION

After analyzing the data using PCA, cluster analysis, and logistic analysis, pitchers seemed to affect a team making the playoffs the most. However, hitting also played a role in determining if a team would get the increased revenue. Hitting itself, will typically not put a team into the playoffs (see Colorado) but it will act as a tie breaker for equal pitching staffs, (see Atlanta11 and Atlanta12). While it may seem obvious pitching and hitting play the largest role in determining success of teams, it is interesting to note the style of play (i.e. “small ball,” “power ball,” etc.) added very little to determining a team’s playoff destiny.

If we were to run this analysis again to improve the model, factors such as team salaries, playoff experience of the roster, average age of the roster, etc. could be included to help make it more accurate. If we decided to go into more detail about the results we found, an in depth analysis of teams pitching staffs could be conducted since pitching is the leading factor. Some possible questions to be investigated are: impact of starting pitchers vs relief pitchers, is it better to have a highly paid star pitcher or a well-rounded pitching rotation, what role do catchers play, or how do pitching coaches effect a team’s staff.

Overall, our method with a 14% misclassification rate adequately measures whether teams make it to the playoffs or not. It could potentially be used to help teams determine if they are going to earn an extra 20 to 30 million dollars in any particular year.

RESOURCES

Barmasse, Jason. "How Much Value Does a Postseason Appearance Hold for MLB Franchises?" *The Fields of Green.* USA Today Sports, 02 Oct. 2014. Web. 02 May 2016. <http://thefieldsofgreen.com/2014/10/02/how-much-value-does-a-postseason-appearance-hold-for-mlb-franchises/>.

"MLB Team Statistics." *ESPN*. Data Provided by Elias Sports Bureau, 2016. Web. <http://espn.go.com/mlb/stats/team/_/stat/batting/year/2016/seasontype/2>.

Appendix A1

Team	Name of MLB team (response variable)
League	categorical variable, either American League and National League
Playoffs	categorical variable, teams that reached or did not reach playoffs
BatAB	at bats
BatR	runs
BatH	hits
Bat2B	doubles
Bat3B	triples
BatHR	home runs
BatTB	total bases; H+ 2B + (3B times 2) + (HR times 3)
BatRBI	runs batted in; A run scored because of the action of a batter
BatAVG	batting average; H divided by AB
BatOBP	on base percentage; (H + BB + HBP) divided by (AB + BB + HBP + SF) (Hits + walks + hit by pitch)/(at-bats + walks + hit by pitch + sacrifice flies)
BatSLG	slugging percentage; TB divided by AB
BatOPS	OPS = OBP +SLG
BatAB/HR	at bats to home run ratio
PitchERA	Earned-run average; (ER times 9) divided by IP (Earned runs*9)/(innings pitched)
PitchSV	saves; Earned when a pitcher finishes a game without having given up the lead after entering in a save situation
PitchCG	complete games; the act of a pitcher pitching an entire game without the benefit of a relief pitcher.
PitchSHO	shutouts, refers to the act by which a single pitcher pitches a complete game and does not allow the opposing team to score a run.
PitchQS	quality starts; pitcher completes at least six innings and permits no more than three earned runs.
PitchR	runs
PitchBB	walks
PitchSO	strikeouts
PitchBAA	opponents batting average; measures a pitcher's ability to prevent hits during official at bats.
PitchSV%	save percentage
PitchK/BB	strikeout to walk ratio
FieldFPCT	fielding percentage (PO + A) divided by (PO + A + E) (Putouts + Assists)/(Putouts + Assists + Errors)

Appendix A2

Correlation Matrix

		BatAB	BatR	BatH	Bat2B	Bat3B	BatHR	BatTB	BatRBI	BatAVG	BatOBP
BatAB	BatAB	1.0000	0.5421	0.7859	0.5659	0.1244	0.2119	0.6393	0.5407	0.6582	0.4992
BatR	BatR	0.5421	1.0000	0.7482	0.6148	0.0836	0.6623	0.9072	0.9959	0.7482	0.8324
BatH	BatH	0.7859	0.7482	1.0000	0.6811	0.2387	0.2549	0.8074	0.7459	0.9824	0.8117
Bat2B	Bat2B	0.5659	0.6148	0.6811	1.0000	0.1125	0.1635	0.6245	0.6103	0.6600	0.6381
Bat3B	Bat3B	0.1244	0.0836	0.2387	0.1125	1.0000	-.2535	0.0978	0.0685	0.2507	0.1209
BatHR	BatHR	0.2119	0.6623	0.2549	0.1635	-.2535	1.0000	0.7566	0.6752	0.2462	0.3524
BatTB	BatTB	0.6393	0.9072	0.8074	0.6245	0.0978	0.7566	1.0000	0.9117	0.7909	0.7523
BatRBI	BatRBI	0.5407	0.9959	0.7459	0.6103	0.0685	0.6752	0.9117	1.0000	0.7461	0.8351
BatAVG	BatAVG	0.6582	0.7482	0.9824	0.6600	0.2507	0.2462	0.7909	0.7461	1.0000	0.8389
BatOBP	BatOBP	0.4992	0.8324	0.8117	0.6381	0.1209	0.3524	0.7523	0.8351	0.8389	1.0000
BatSLG	BatSLG	0.4984	0.9019	0.7333	0.5764	0.0815	0.8081	0.9851	0.9073	0.7428	0.7360
BatOPS	BatOPS	0.5301	0.9352	0.8063	0.6344	0.0999	0.7029	0.9675	0.9400	0.8225	0.8731
BatAB_HR	BatAB/HR	-.1497	-.6028	-.1958	-.1375	0.2200	-.9582	-.6999	-.6158	-.1924	-.3200
PitchERA	PitchERA	0.2391	0.0604	0.1932	0.1360	0.1264	0.0538	0.1690	0.0498	0.1631	-.0378
PitchSV	PitchSV	-.1054	0.0233	-.0409	-.0282	-.0919	0.0957	0.0241	0.0334	-.0170	0.0607
PitchCG	PitchCG	-.1018	0.1039	0.0118	0.0120	-.0307	0.0106	0.0110	0.1104	0.0468	0.1800
PitchSHO	PitchSHO	-.1552	-.0006	-.0424	-.1039	-.0393	-.0598	-.0807	0.0011	-.0014	0.1581
PitchQS	PitchQS	-.0865	0.1289	0.0059	0.0047	0.0900	0.1514	0.1098	0.1271	0.0312	0.0673
PitchR	PitchR	0.2585	0.0457	0.1866	0.1481	0.1201	0.0281	0.1507	0.0346	0.1495	-.0372
PitchBB	PitchBB	0.1163	-.1413	0.0256	0.0215	0.1361	-.1308	-.0499	-.1501	-.0038	-.1518
PitchSO	PitchSO	-.0178	0.0173	-.0400	-.0395	-.1465	0.1228	0.0334	0.0351	-.0429	0.0704
PitchBAA	PitchBAA	0.2216	-.0066	0.1948	0.1075	0.1688	-.0853	0.0825	-.0153	0.1708	-.0356
PitchSV_	PitchSV%	-.1905	0.0059	-.1028	-.1594	-.0297	-.0160	-.0917	0.0163	-.0682	0.0426
PitchK_BB	PitchK/BB	-.1005	0.1035	-.0462	-.0391	-.1803	0.1635	0.0519	0.1172	-.0261	0.1425
FieldFPCT	FieldFPCT	0.0594	0.0968	0.0268	-.0431	-.2319	0.1972	0.1036	0.1237	0.0166	0.0849

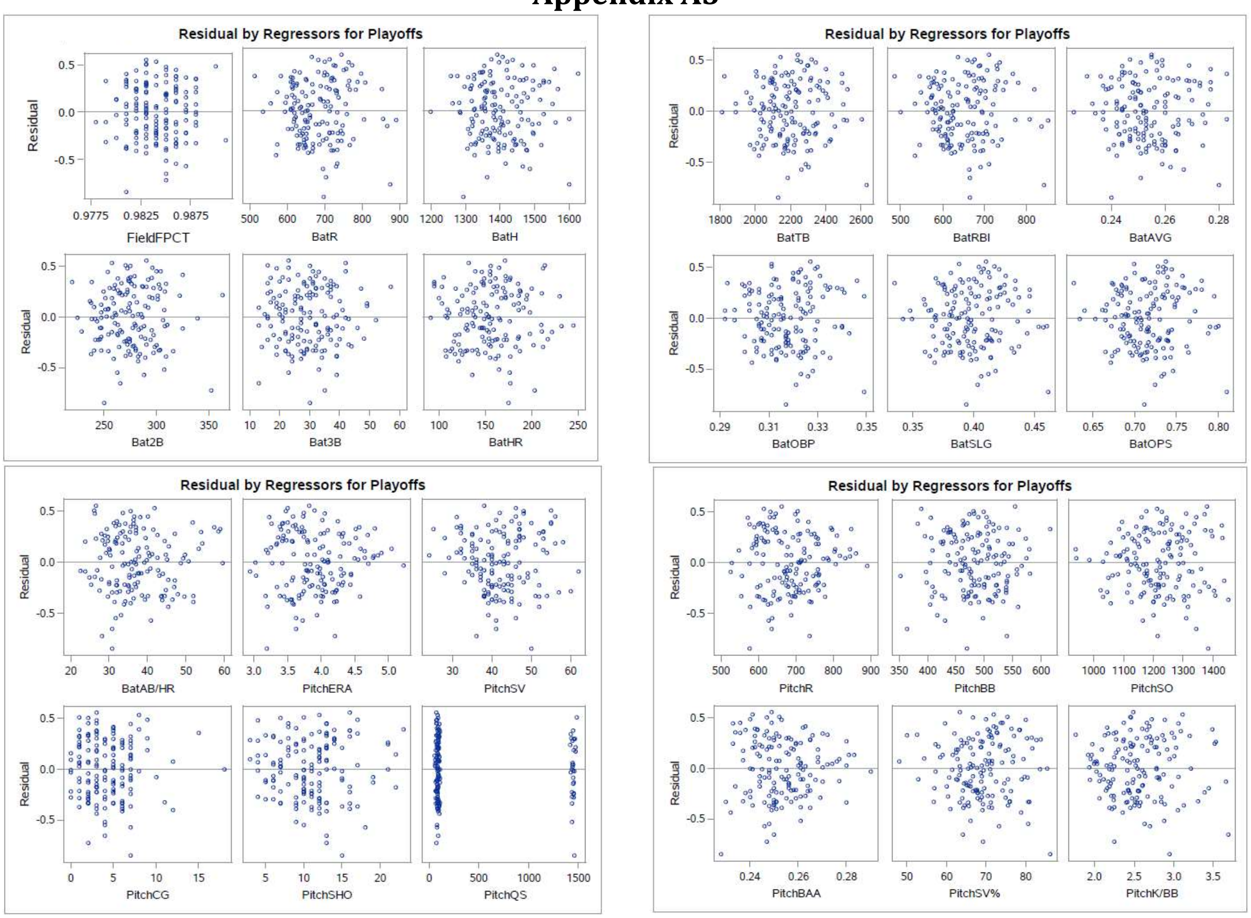
Correlation Matrix

	BatSLG	BatOPS	BatAB_HR	PitchERA	PitchSV	PitchCG	PitchSHO	PitchQS	PitchR	PitchBB
BatAB	0.4984	0.5301	-.1497	0.2391	-.1054	-.1018	-.1552	-.0865	0.2585	0.1163
BatR	0.9019	0.9352	-.6028	0.0604	0.0233	0.1039	-.0006	0.1289	0.0457	-.1413
BatH	0.7333	0.8063	-.1958	0.1932	-.0409	0.0118	-.0424	0.0059	0.1866	0.0256
Bat2B	0.5764	0.6344	-.1375	0.1360	-.0282	0.0120	-.1039	0.0047	0.1481	0.0215
Bat3B	0.0815	0.0999	0.2200	0.1264	-.0919	-.0307	-.0393	0.0900	0.1201	0.1361
BatHR	0.8081	0.7029	-.9582	0.0538	0.0957	0.0106	-.0598	0.1514	0.0281	-.1308
BatTB	0.9851	0.9675	-.6999	0.1690	0.0241	0.0110	-.0807	0.1098	0.1507	-.0499
BatRBI	0.9073	0.9400	-.6158	0.0498	0.0334	0.1104	0.0011	0.1271	0.0346	-.1501
BatAVG	0.7428	0.8225	-.1924	0.1631	-.0170	0.0468	-.0014	0.0312	0.1495	-.0038
BatOBP	0.7360	0.8731	-.3200	-.0378	0.0607	0.1800	0.1581	0.0673	-.0372	-.1518
BatSLG	1.0000	0.9725	-.7584	0.1357	0.0522	0.0354	-.0548	0.1433	0.1111	-.0873
BatOPS	0.9725	1.0000	-.6560	0.0851	0.0601	0.0855	0.0127	0.1250	0.0673	-.1130
BatAB_HR	-.7584	-.6560	1.0000	-.0355	-.0777	-.0375	0.0467	-.1503	-.0133	0.1492
PitchERA	0.1357	0.0851	-.0355	1.0000	-.5982	-.2381	-.5846	0.1065	0.9863	0.5298
PitchSV	0.0522	0.0601	-.0777	-.5982	1.0000	-.0732	0.3310	-.0015	-.6001	-.3518
PitchCG	0.0354	0.0855	-.0375	-.2381	-.0732	1.0000	0.3100	-.0023	-.2436	-.3256
PitchSHO	-.0548	0.0127	0.0467	-.5846	0.3310	0.3100	1.0000	-.0784	-.5990	-.3854
PitchQS	0.1433	0.1250	-.1503	0.1065	-.0015	-.0023	-.0784	1.0000	0.1034	0.0669
PitchR	0.1111	0.0673	-.0133	0.9863	-.6001	-.2436	-.5990	0.1034	1.0000	0.5286
PitchBB	-.0873	-.1130	0.1492	0.5298	-.3518	-.3256	-.3854	0.0669	0.5286	1.0000
PitchSO	0.0442	0.0556	-.1542	-.5492	0.3036	0.0419	0.3813	-.0049	-.5438	-.2463
PitchBAA	0.0428	0.0191	0.0970	0.8643	-.4363	-.2672	-.5719	0.0228	0.8692	0.3480
PitchSV_	-.0620	-.0286	0.0175	-.5452	0.7012	0.1121	0.3328	0.0791	-.5503	-.3522
PitchK_BB	0.0860	0.1093	-.1951	-.6570	0.3876	0.2467	0.4786	-.0474	-.6530	-.8226
FieldFPCT	0.1038	0.1044	-.1679	-.2490	0.2159	0.1476	0.1607	-.1249	-.3306	-.2663

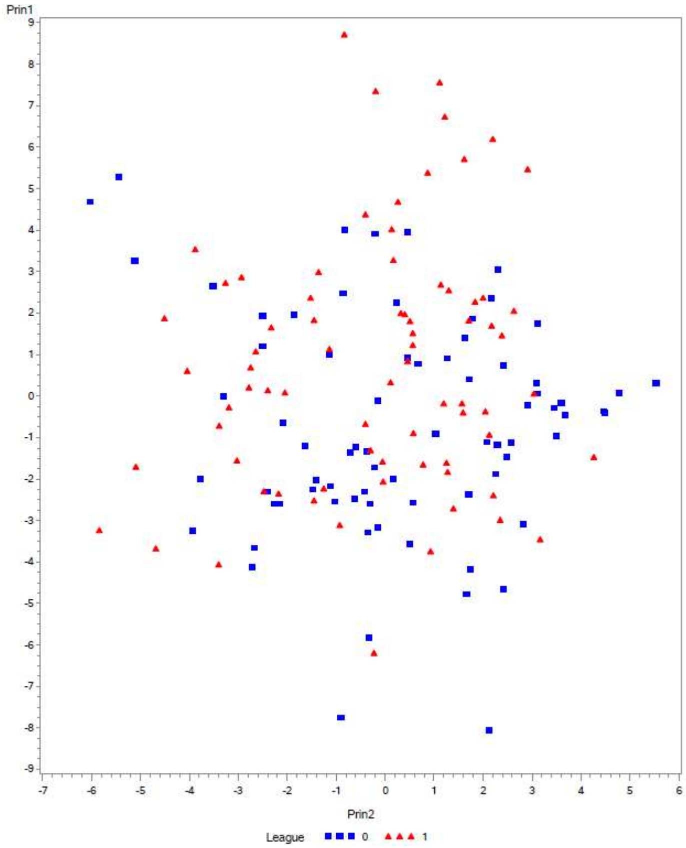
Correlation Matrix

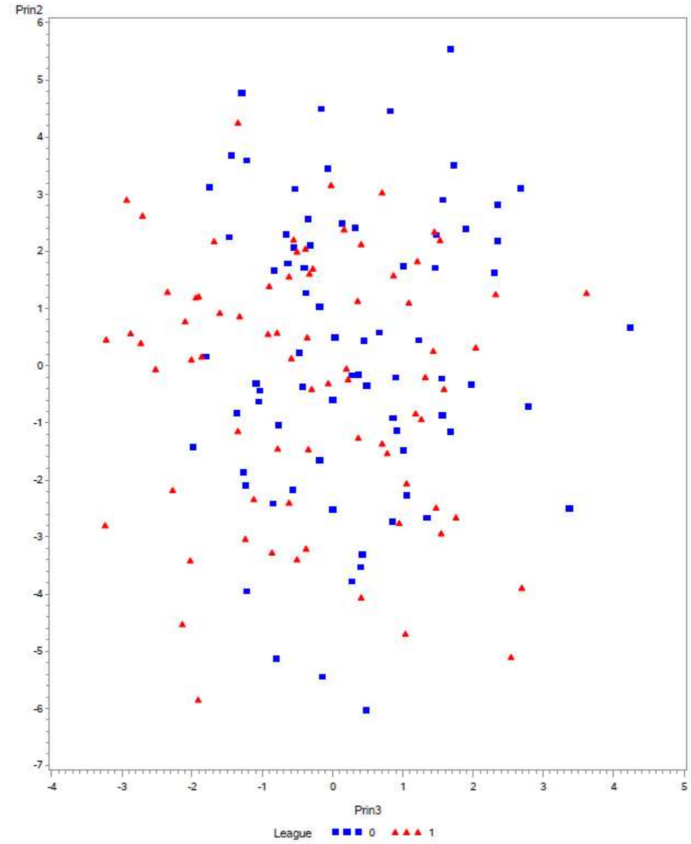
	PitchSO	PitchBAA	PitchSV_	PitchK_BB	FieldFPCT
BatAB	-.0178	0.2216	-.1905	-.1005	0.0594
BatR	0.0173	-.0066	0.0059	0.1035	0.0968
BatH	-.0400	0.1948	-.1028	-.0462	0.0268
Bat2B	-.0395	0.1075	-.1594	-.0391	-.0431
Bat3B	-.1465	0.1688	-.0297	-.1803	-.2319
BatHR	0.1228	-.0853	-.0160	0.1635	0.1972
BatTB	0.0334	0.0825	-.0917	0.0519	0.1036
BatRBI	0.0351	-.0153	0.0163	0.1172	0.1237
BatAVG	-.0429	0.1708	-.0682	-.0261	0.0166
BatOBP	0.0704	-.0356	0.0426	0.1425	0.0849
BatSLG	0.0442	0.0428	-.0620	0.0860	0.1038
BatOPS	0.0556	0.0191	-.0286	0.1093	0.1044
BatAB_HR	-.1542	0.0970	0.0175	-.1951	-.1679
PitchERA	-.5492	0.8643	-.5452	-.6570	-.2490
PitchSV	0.3036	-.4363	0.7012	0.3876	0.2159
PitchCG	0.0419	-.2672	0.1121	0.2467	0.1476
PitchSHO	0.3813	-.5719	0.3328	0.4786	0.1607
PitchQS	-.0049	0.0228	0.0791	-.0474	-.1249
PitchR	-.5438	0.8692	-.5503	-.6530	-.3306
PitchBB	-.2463	0.3480	-.3522	-.8226	-.2663
PitchSO	1.0000	-.6291	0.2994	0.7387	0.1873
PitchBAA	-.6291	1.0000	-.4494	-.5810	-.2693
PitchSV_	0.2994	-.4494	1.0000	0.3797	0.2262
PitchK_BB	0.7387	-.5810	0.3797	1.0000	0.2705
FieldFPCT	0.1873	-.2693	0.2262	0.2705	1.0000

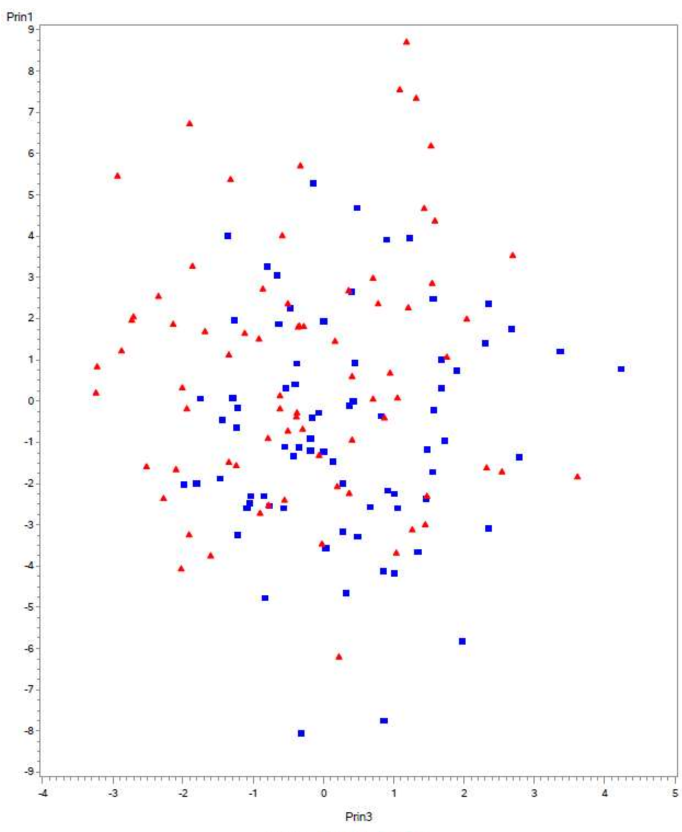
Appendix A3

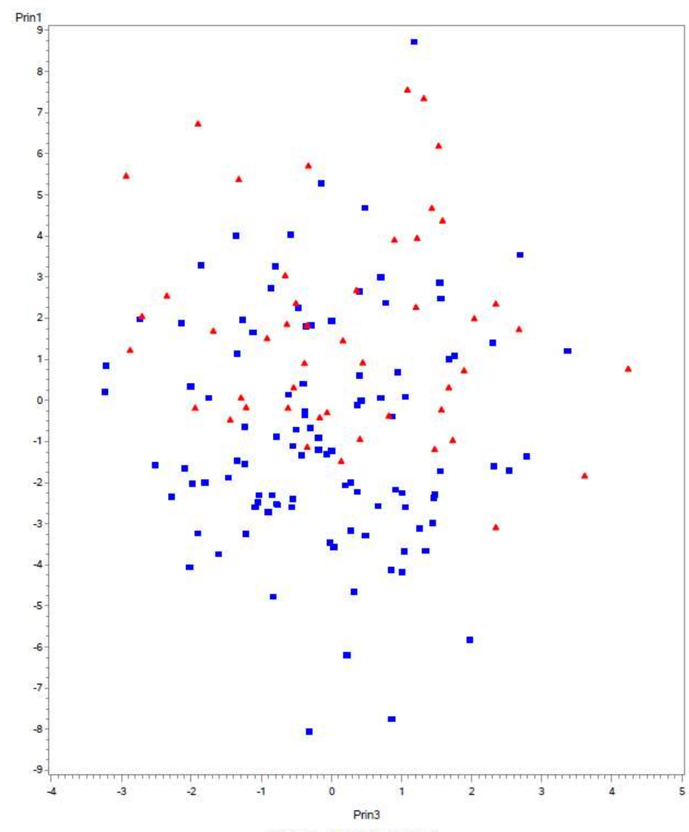


Appendix A4

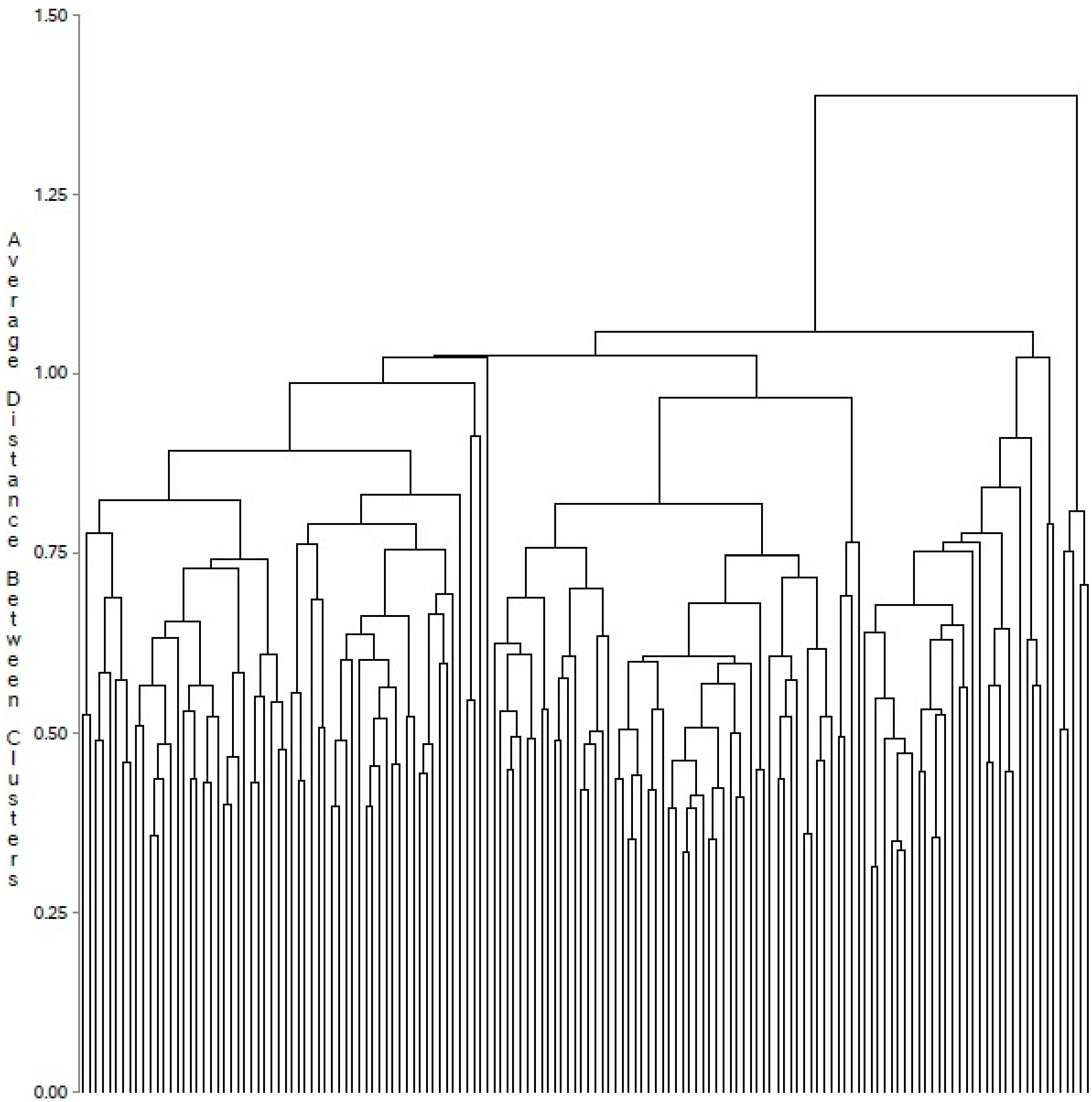

Prin1
Prin2
League ■■■ 0 ▲▲▲ 1


Prin2
Prin3
League ■■■ 0 ▲▲▲ 1


Prin1
Prin3
League ■■■ 0 ▲▲▲ 1


Prin1
Prin3
Playoffs ■■■ 0 ▲▲▲ 1

Appendix A5


1.50
1.25
1.00
0.75
0.50
0.25
0.00
Average Distance Between Clusters