By the Numbers

Volume 28, Number 1 | The Newsletter of the SABR Statistical Analysis Committee | April, 2018

Review

Academic Research: New and Old

Charlie Pavitt

The author reviews three academic papers: one examining player performance trends after a team change; one adding to sabermetricians' previous research on count values; and an 80-year-old study examining home field advantage.

Bryan L. Rogers, James M. Vardaman, David G. Allen, Ivan S. Muslin, and Meagan Brock Baskin (2017), Turning up by turning over: The change of scenery effect in major league baseball, Journal of Business and Psychology, Vol. 32, pp. 547-560.

This is an example of a potentially interesting study that is compromised due to insufficient thinking about the relevant issue.

The authors compared position players who switched to another team after two seasons with at least 100 PA for the same team (2004-2015). They compared players who had been declining across the past two years (422 cases) versus those who were stable or improving (290 cases).

In this issue

The previous issue of this publication was July, 2017 (Volume 27, Number 1).

Performance indices included BA, OPS, wRC+, and fielding percentage, with a set of controls including (among others) career batting average and age "to account for the effects of a player's declining ability;" I think it would have been better if the authors had computed a career trajectory and then used the relevant season's estimate from that trajectory as a stand-in for career BA.

There were also comparisons to players with declines (922 cases) versus stable or improving performance who did not change teams.

The findings:

1. On average, those who declined across years one and two improved in year three on all three batting indices, a simple case of regression to the mean.

Those who switched teams declined more than those who remained, which I hypothesize indicates that those who declined more were more likely to switch teams than those who stayed; the authors do not seem to have performed this easy comparison. It did not matter whether they switched through trades or free agency.

2. On average, those who were stable or improved across years one and two declined in year three on all three batting indices, again more so if they changed teams than if they did not, with the same implications as beforehand. Again, trades or free agency did not matter.

3. These regressions to the mean remained stable for year four for those remaining with the year-three team.

4. Fielding percentage changed in unison with the batting indices, but was never statistically significant. Rogers and associates realized that the issue here was the absence of variance in fielding percentage; obviously, they should have used one of the range-factor-type measures instead.

Harold M. Cavins (1938), A study to discover the relative numbers of baseball games won at home and away from home in the major leagues, Research Quarterly, Vol. 8, pp. 57-59.

My summaries are usually of recent work, but I thought there might be interest in an 80-year-old study I recently learned of on home field advantage based on 1932-1937 team performance. Overall, the American League percentage of home team wins was .5474; the National League .5526. Among the sixteen teams, the range was from .527 (St. Louis Browns) to .573 (Cleveland Indians). Of the 96 total cases of each team for each season, 11 were greater than .600 and eight less than .500 (the lowest was .460). Second-division teams seemed to benefit more than first-division (.5601 versus .5170) during the first five years of data, but this tendency reversed in the final sixth season.

This is not *the* earliest serious analytic work -- credit F. C. Lane's 1917 prophetically accurate studies of the relative run value of different plate appearance outcomes for that -- but it is certainly one of the earliest. To the best of my knowledge, about forty years passed before the next serious study of this topic. The overall 55 percent advantage across season and league in Cavins' data is close to current 53-54 percent estimates.

Philippa Swartz, Mike Grosskopf, Derek Bingham, and Tim B. Swartz (2017), The quality of pitches in major league baseball, The American Statistician, Vol. 71 No. 2, pp. 148-154.

Over the years there has been quite a bit of work on the impact of the count on the outcome of the plate appearance. Some of it looks at the outcomes of plate appearances when they end at particular counts (i.e., what batter performance is when the ball is put into play with, say, a 1-0 versus 0-1), and some at the outcomes of plate appearances when they are passed through along the way (i.e., what batter performance is when the ball is put play at some point when, earlier on, the count was 1-0 versus 0-1). There is even a study of the run value of a ball and a strike at different counts.[1]

The authors, based on more than two million pitches PITCHf/x reported between 2013 and 2015, presented something along the lines of Walsh's work; the number of expected bases for batters in each count passed along the way given that the next pitch was either a strike or ball (in other words not put into play):

Count	0-0	0-1	0-2	1-0	1-1	1-2	2-0	2-1	2-2	3-0	3-1	3-2
Ball	.51	.42	.32	.61	.51	.39	.79	.69	.55	1.00	1.00	1.00
Strike	.37	.28	.00	.42	.31	.00	.51	.38	.00	.68	.54	.00

There are two types of obvious cases – a fourth ball has to be 1.00 and a third strike (assuming no passed ball) has to be .00. Otherwise, note that in every case in which the result of the pitch was the same count, the figures are either exactly the same (such as a ball with an 0-1 count and a strike with a 1-0 count, both making it 1-1; .42 in this case) or differing by .01 (a ball at 0-2 is .32 and a strike at 1-1 is .31). This means that plate appearance outcomes for counts are unaffected by the sequence of balls and strikes passed through to get there.

The authors used this data to argue that starting pitchers are most effective keeping expected runs low between pitches 20 and 70; needing to warm up a bit with the first 20 pitches after the game starts, becoming somewhat less effective between pitches 80 and 100, and far less so after 100. However, they didn't consider that the first 20 pitches are always to the top of the order, thus probably accounting for the higher scoring seen there.

Swartz et al. also presented a table of the best pitchers in this regard during the 2013-2015 interim, some of whom were no surprise to me (Jordan Zimmerman, Max Scherzer, Clayton Kershaw, David Price, Chris Sale, Madison Bumgarner) and some of whom were (Phil Hughes, Vidal Nuño, Kevin Slowey, Hisashi Iwakuma).

Charlie Pavitt, chazzq@udel.edu ♦

[1] See John Walsh's piece online at The Hardball Times website: http://www.hardballtimes.com/main/article/searching-for-the-games-best-pitch/

Does Lineup Balance Matter?

Shlok Goyal

Baseball theory says that having a balance of contact and power hitters in a lineup improves a team's run production. After conducting an analysis determining the correlation between a team's runs scored and its lineup balance, the author finds that lineup balance does matter, but not much: teams would gain at best a negligible number of runs by balancing their power hitters with contact hitters.

Conventional wisdom has long held that a lineup maximizes its run-producing ability when it has a mix of contact hitters and sluggers working in tandem. The sluggers will then have someone to drive home with an extra-base hit; whereas, on a team composed entirely of power hitters, such a hit is likely to come with nobody on base. On the flip side, a team full of contact hitters is wont to leave many men on base since it requires a string of hits to score a run. Thus, an imbalanced lineup should score fewer runs than its hitters' overall caliber would indicate.

To test this hypothesis, we first need to create a statistic that will give us a rough measure of the balance of a lineup. We can do this by measuring how slugger-heavy (or contact-heavy) a lineup is. Since sluggers hit more doubles and home runs than the average, we can measure slugger-heaviness by calculating what proportion of a team's runs comes from doubles and home runs relative to the total runs created by the team's hits. Using the weights from weighted On-Base Average (wOBA)—a statistic that assigns each plate appearance a value of how many runs it is expected to create based on its outcome—we can measure a team's propensity to slug. Call this statistic Power Propensity Index (PPI):

$$PPI = \frac{1.242*2B + 2.015*HR}{0.878*1B + 1.242*2B + 1.569*3B + 2.015*HR}*100\%$$

The coefficients are from wOBA's weights for each type of hit in 2016, as given on FanGraphs. Note that even though teams that slug more tend to walk more and strike out more, walks and strikeouts don't determine whether a hitter is a slugger or a contact hitter and so are not included in the equation.

We can rank teams based on this metric to identify how balanced their lineup was.[1] See Table 1. In the table, teams in the middle had the most balanced lineups, while those at the top tended to focus on hitting extra-base hits and those at the bottom on hitting singles.

The results largely fall in line with what one would expect. The 2016 Orioles and Rays had offenses built almost entirely on power—respectively, 17.9% and 16.2% of each team's hits were home runs. Meanwhile, the Giants and Marlins had mostly contact hitters on their teams—Brandon Belt led the Giants with 17 home runs, while even Giancarlo Stanton couldn't change the fact that only 8.8% of the Marlins' hits were dingers.

Now, we need to measure the impact of an imbalanced lineup on runs scored. Since a real game is affected by numerous factors such as base-running ability and opponents' defensive abilities, not to mention plain randomness, we evaluate lineups via Tom Tango's Markov simulation, which predicts the runs per game a lineup of hitters with given batting statistics would score.[2]

[1] Our main goal is reasonably measuring a lineup's tendency to hit for power, which lets us know how balanced it is between hitting for power and hitting for average. Any statistic that does this will give results similar to those given by PPI. For example, it is possible to use Isolated Power (ISO) for this metric. The one small difference will be that triples are treated as a sign of power in ISO, whereas sluggers rarely hit triples, but the results will largely be the same.

[2] Tango's Markov simulation can be found here: http://www.tangotiger.net/markov.html

I followed these steps to find the impact each percentage change in PPI has on runs scored:

First, I divided the original batting line into two -- one with a higher PPI and one with a lower PPI. The two summed to the same values as the original line. (You can think of each of the two batting lines as a half-season's worth of statistics, one that is more oriented towards contact hitting and one more towards power hitting.)

Baserunning skill was considered to be identical between the two lines. I assigned strikeouts and walks so that the line with the higher PPI would have more of both. The exact number of strikeouts and walks was determined by regression equations fitting PPI vs. strikeouts and walks.

Both batting lines were created to have the same number of batting outs (AB-H), since we want the half-seasons to represent the same number of innings played. Equally important, the two batting lines were selected to be as close as possible to each other in expected runs scored per game (R/G), as evaluated by the Markov simulation.

Second, I averaged the runs scored per game of the two batting lines. I calculated the difference between this average and the expected runs/game of the original batting line. This gives us the difference in runs/game caused by lineup imbalance.

This gives us the run impact of the difference of the imbalance, but not the difference itself. To find that, I took the difference between the original PPI and the lower PPI as well as between the original PPI and the higher PPI. Then, I averaged the absolute values of those two figures.

Finally, I repeated this process for many teams, and plotted the data to figure out the relationship.

By performing these steps, we first took two imbalanced lineups that should have, on average, performed the same as the original.

Comparing the average R/G they yield with the original lineup's R/G, then, would let us assess the impact of changing PPI. As this PPI differs from the original PPI to a greater and greater extent, the difference in R/G between the original and the imbalanced lineups enables us to calculate how many runs a team loses as it becomes more and more imbalanced.

Table 1

Team	PPI
Orioles	51.5%
Rays	51.4%
Cardinals	50.7%
Blue Jays	50.6%
Astros	49.0%
Mets	48.3%
Cubs	48.1%
Brewers	47.8%
Twins	47.7%
Red Sox	47.4%
Padres	47.2%
Indians	47.1%
Nationals	46.9%
Mariners	46.8%
Dodgers	46.8%
Rockies	46.4%
Rangers	46.3%
Athletics	45.3%
Tigers	44.8%
Diamondbacks	44.5%
Yankees	44.3%
Reds	43.7%
White Sox	43.5%
Angels	43.2%
Phillies	42.5%
Pirates	42.1%
Braves	40.9%
Royals	40.0%
Giants	39.4%
Marlins	37.4%

To better see how all this works, let's take an example. Assume the 2016 Tigers' batting line is our original data:

Season	Team	AB	H	2B	3B	HR	BB	SO	PPI	Markov R/G
2016	Tigers	5526	1476	252	30	211	493	1303	44.8%	4.944

Then, we split this line into two parts, one with a higher PPI than the original and one with a lower PPI.

Season	Team	AB	H	2B	3B	HR	BB	SO	PPI	Markov R/G
2016	Total	5526	1476	252	30	211	493	1303	44.8%	4.944
2016	High PPI 1	2695	670	166	13	118	278	720	56.0%	4.939
2016	Low PPI 1	2831	806	86	17	93	215	583	34.4%	4.940

Averaging High PPI 1 and Low PPI 1's expected R/G yields 4.9395, which is lower than the original line's R/G by 0.0045. High PPI 1 differs from the original by 11.2% while Low PPI 1 differs by 10.4%. Averaging the two yields a lineup imbalance of 10.8%. Thus, at a lineup imbalance of 10.8%, we can expect the team to underperform by 0.0045 runs/game (0.729 runs/season).

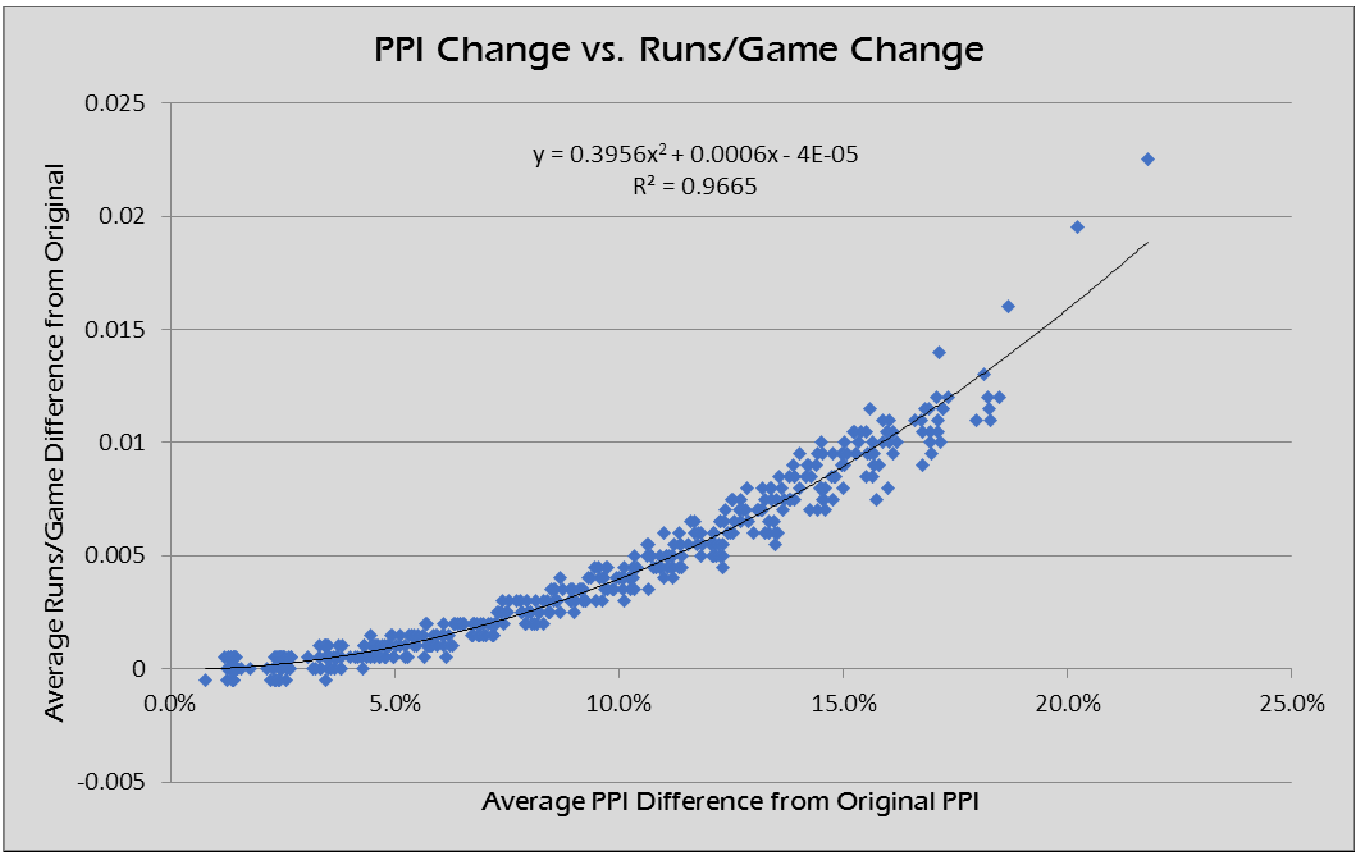
We can try again by splitting the original line into two in a different way, like this:

Season	Team	AB	H	2B	3B	HR	BB	SO	PPI	Markov R/G
2016	Total	5526	1476	252	30	211	493	1303	44.8%	4.944
2016	High PPI 2	2676	651	178	13	121	287	741	59.4%	4.937
2016	Low PPI 2	2850	825	74	17	90	206	562	31.5%	4.936

We go through the same calculations here. Averaging High PPI 2 and Low PPI 2's expected R/G gives 4.9365, lower than the original line's by 0.0075. Averaging the difference between High PPI 2's PPI and the original line's PPI and High PPI 1's PPI and the original line's PPI gives 13.95%. So if a lineup is imbalanced by 13.95% PPI, the team will underperform by 0.0075 runs/game (1.215 runs/season).

I performed the above steps using as my data the batting statistics for each of the 30 teams in the 2016 season (although the analysis can be conducted with any batting line). I split each of the 30 teams into 15 sets of two batting lines in the same fashion as described above. Each set had a different difference in PPI to measure how runs/game changes as PPI gets further away from the original.

Plotting the resulting 450 datapoints demonstrates a fairly clear positive exponential relationship:



As the difference in PPI rises (as the lineup becomes more imbalanced), a team scores increasingly fewer runs than its batting statistics would indicate. This underperformance increases exponentially with each change in PPI since the effects of lineup imbalance would multiply on itself as a team becomes too focused on slugging or on hitting for contact—after a while, every home run will become a solo shot or three singles always required to drive home a run.

If the original lineup was at an ideal PPI, both its "children" lineups would be slightly more imbalanced. Thus, they would both score slightly fewer runs than the original. If the original itself was at an extreme PPI, then one of its children lineups will actually perform better than the original, but due to the exponential nature of the curve, the underperformance of the other child lineup will cause the average of the two to underperform.

The graph helps us understand the impact of lineup imbalance on runs lost per game. It does not necessarily tell us how many runs/game a team would gain or lose by changing its PPI in one direction. This is because as teams move closer to the ideal PPI, they will perform better and as they move away from it, they will perform worse.

Looking at the graph, it seems that the theory valuing a balance of contact and power hitters is justified. However, this balance doesn't make a significant difference in runs scored. Consider, for example, that the difference between the team with the highest and lowest PPI in 2016 was 14.1%. Even if we go out on a limb and assume that one of these teams is at the ideal PPI, the increase in runs scored for the other team in changing its PPI level by 14.1% would be a paltry 0.008 runs/game or 1.29 runs/season. That's a remarkably small improvement for such a dramatic change in a lineup's composition.

Furthermore, a team cannot precisely predict how its players will hit. A player with a PPI of 40% one year could easily be at 45% the next just by luck. Similarly, a team's PPI could change by a few percentage points each season with the same personnel. It makes it nearly impossible then even to salvage that additional one run over the course of a season. This result also assumes that the hitters' talent level remains the same. With the limited supply of players talented enough to play in the majors, it is difficult for a team to find players with equal skill but different PPI, further lowering the chances of saving that one extra run.

Unless every hitter on your team is Joey Gallo, there is little to be gained by trying to balance your lineup to include both contact and power hitters. In that light, teams would be better off finding the best players they could without worrying about how that player fits into their lineup. There is one caveat to this. The analysis assumes that base-running and basestealing skills are the same among every team. Teams with a more contact-oriented approach, however, might prefer speedier players who can get into scoring position more often.

Baseball wisdom is right and wrong. Yes, having a balanced lineup increases the number of runs scored. In most cases, however, the effect is so minimal as to be negligible. Pick the best hitters you can; they will work together just fine, irrespective of lineup balance.

Shlok Goyal, shlok.goyal@gmail.com ♦

Data-Driven Clustering of Recent World Series Champions

Mark Kindem

How have recent World Series teams been built? Teams have varied in the extent that they have relied on free agency, drafting, trades, player development, and international scouting. With so many parameters, it's hard to get an intuitive grasp of which teams are which. In this study, the author uses the technique of "statistical clustering" to group like teams together, creating a manageable number of (generally similar) approaches to team-building that can be examined individually.

1. Introduction

While this article is not necessarily about the 2017 Houston Astros, their success is what motivated it. After a well-publicized Sports Illustrated cover story in 2014, "tanking" became a real phenomenon in Major League Baseball.[1] The practice is controversial to say the least, especially amongst fans. However, the Astros won the World Series in 2017, exactly as Sports Illustrated predicted they would, so now the perception is that tanking is a viable strategy for building a championship team. However, did the Astros have to forfeit the 2011-2013 seasons to win the World Series in 2017? Or was the strategy described by the Sports Illustrated article merely a myth-building cover-up, disguising a more conventional (and possibly less intentional) general management approach?

This article attempts to answer this in a holistic way, by exploring how each World Series Championship team from the past forty years was constructed, and then using advanced analytics techniques to categorize each winner into a data-driven cluster. The clusters, formed without human interpretation by a statistical computing algorithm, can then be defined by their unique properties, and each cluster will effectively represent a viable strategy for winning the World Series.

2. Methods

Data Collection

The goal of the analysis originally was to cluster each World Series Championship team according to *how many players* on the World Series roster were acquired by these six different acquisition types:

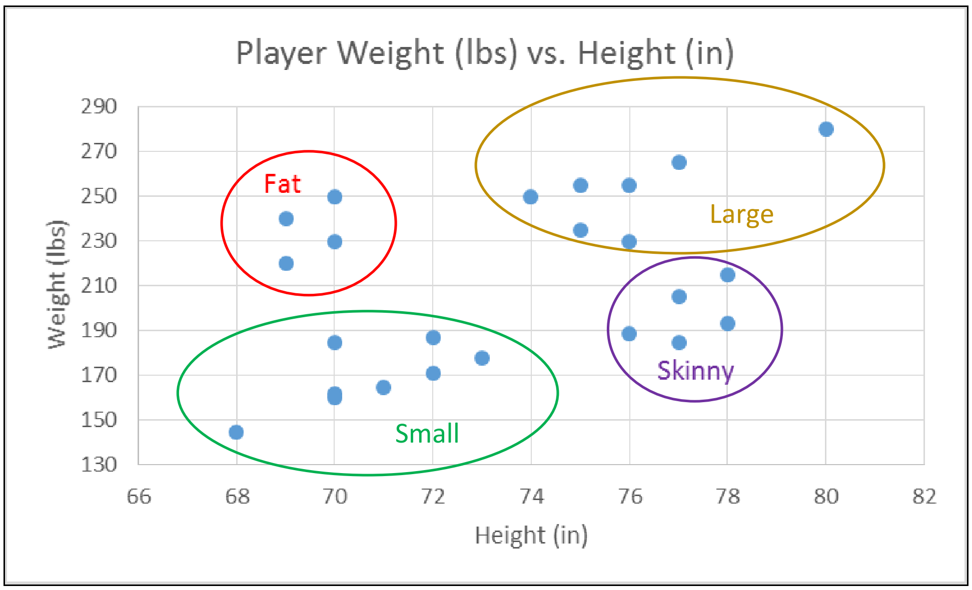
- Draft
- Free Agency
- Non-Season Trade (Player had major league experience at time of trade)
- Non-Season Trade – Minor League (Player did not have major league experience at time of trade)
- Within-Season Trade
- International/Amateur Free Agent Signing

[1] Tanking is a team-building strategy, usually employed by a team on the decline, where a team will rid itself of all of its remaining good players in order to shed payroll and lose lots of games, ostensibly to increase its chances of earning top draft picks.

However, not all players are created equal. Therefore, I decided to collect end-of-season WAR estimates for each player, sum up all WAR estimates for each team, and then determine the percentage of a team's total WAR that came from each of the six acquisition types instead.[2] This idea was expanded to determine the percentage of a team's position player WAR that came from each of the six types, and the percentage of a team's pitching WAR that came from each of the six types. This created 18 variables for each team, and these variables are the basis of the subsequent clustering. WAR was chosen as the measure of player quality since it captures all player performance in one statistic and allows for valid comparisons between hitters and pitchers.

Statistical Clustering

Statistical clustering is a classification exercise that partitions observations into groups with common, similar properties.[3] It is not to be confused with *segmentation*; segmentation uses subjective, person-defined rules to classify observations, while the classification rules used in clustering are defined completely by the data itself. Clustering cases is done statistically by minimizing distance between all observations based on similarity of given variables of interest. To illustrate, consider this simple example. Assume a manager wants to cluster his players on two factors, height and weight. He can visualize his four clusters with a scatterplot:



Statistical clustering is an extension of this two dimensional example applied to n dimensions (in this article, n = 18, since there are 18 factors considered in the clustering algorithm). Conceptually, clustering is straightforward, but what makes it seem difficult is that it is not easy for a person to visualize an eighteen dimensional scatterplot. Fortunately, statistical software packages such as R or SAS can do the underlying mathematics and "see" the clusters for us.

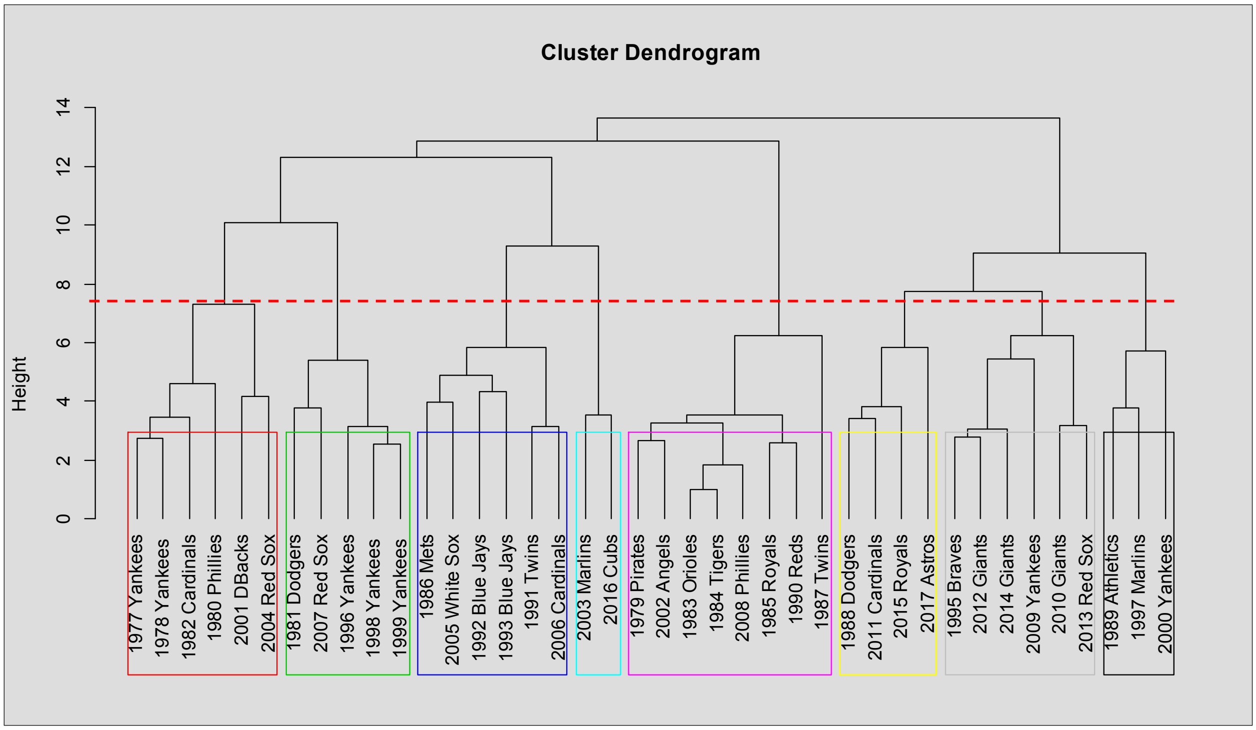
There are two primary clustering methods: Hierarchical Clustering, and Partitive (or K-Means) Clustering. Hierarchical Clustering was chosen for this study, because it works well with smaller datasets, and lends itself to better visualization of the results.

Specifically, Hierarchical Clustering allows results to be seen in a chart called a *dendrogram*. Table 1 shows the dendrogram for the teams in this study:

[2] WAR values were taken from Baseball Reference. An appendix containing full details of the methodology described in this paper is available by e-mailing the author.

[3] For a description of statistical clustering, see https://support.sas.com/rnd/app/stat/procedures/ClusterAnalysis.html .

Table 1: World Series Champion Clusters



The dendrogram can be interpreted in the following way: the similarity between two teams is measured by the height of the lowest horizontal line that connects them.[4] For example, the 1983 Orioles and the 1984 Tigers are the two most similar teams (in terms of their construction), since the height of their horizontal connection is less than 1. The 1983 Orioles are very different from the 1988 Dodgers though, since the height of their connection is 14. In fact, despite their adjacent placement along the spectrum of teams, the 1988 Dodgers and 1987 Twins are very different as well (their connection's height is also 14). Therefore, it is important to remember that relative proximity does not necessarily indicate similarity.

The horizontal hash line is a cutoff; all linked connections beneath this line are aggregated into one cluster. So, by counting the number of times the cutoff intersects with a vertical line, you can count the number of clusters (in this case there are eight). The height of the cutoff may be subjective, but there are analytical techniques that can determine the optimal number of clusters for you.[5] Determining the optimal number of clusters is important though, since we always want to ensure that the similarity of teams within a cluster is maximized, while the total number of clusters is minimized.

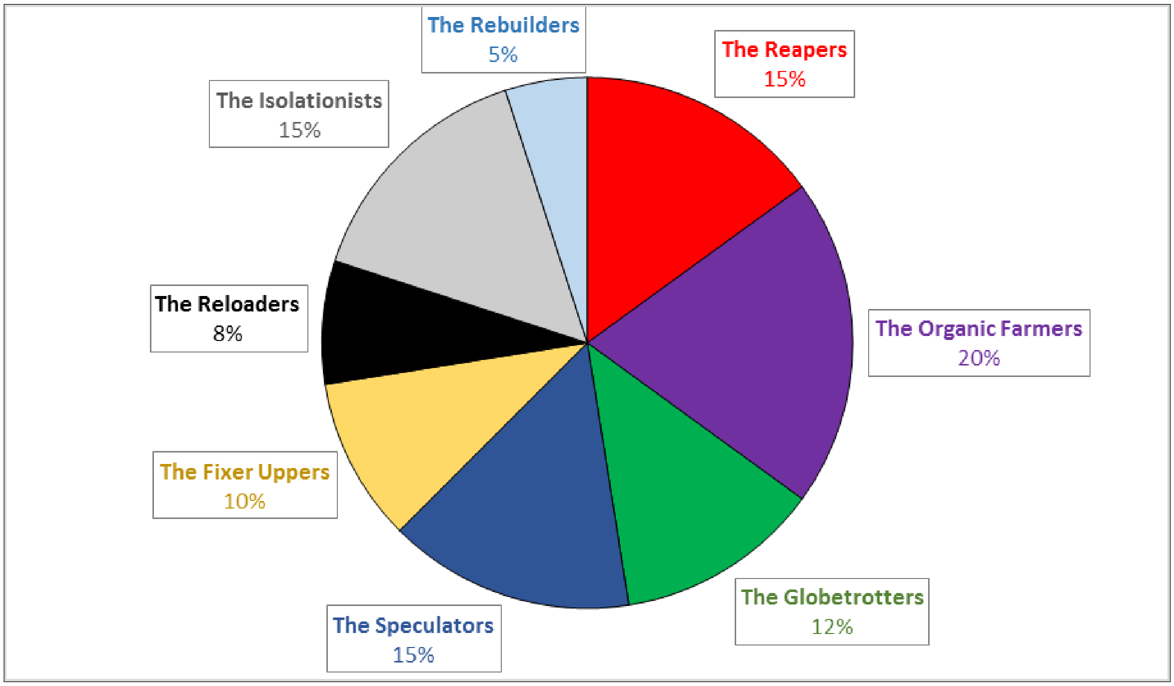
[4] See http://uc-r.github.io/hc_clustering

[5] An appendix containing an explanation of why this line was placed here, and other mathematical decision specifics related to these results, can be obtained from the author.

3. Summary of Clusters

Before exploring each cluster in detail, here's a summary of what the algorithm came up with. First, a pie chart:

Table 2: The Eight Clusters



While the algorithm will create the clusters, it will not "explain" why it chose those clusters, or what the clusters represent. I examined the clusters, tried to determine what teams in each cluster had in common, and chose names for them. The names of the clusters are my creative invention, designed to memorably describe what makes each cluster distinct. However, to reiterate, the clusters' names are their only subjective property. The clusters, and teams assigned to them, are completely determined by the clustering algorithm described in the previous section, and the subsequent analysis is a summary of the resultant output. These clusters may not be the only way to categorize these teams, but the method presented here does have an advantage: the removal of qualitative assumptions

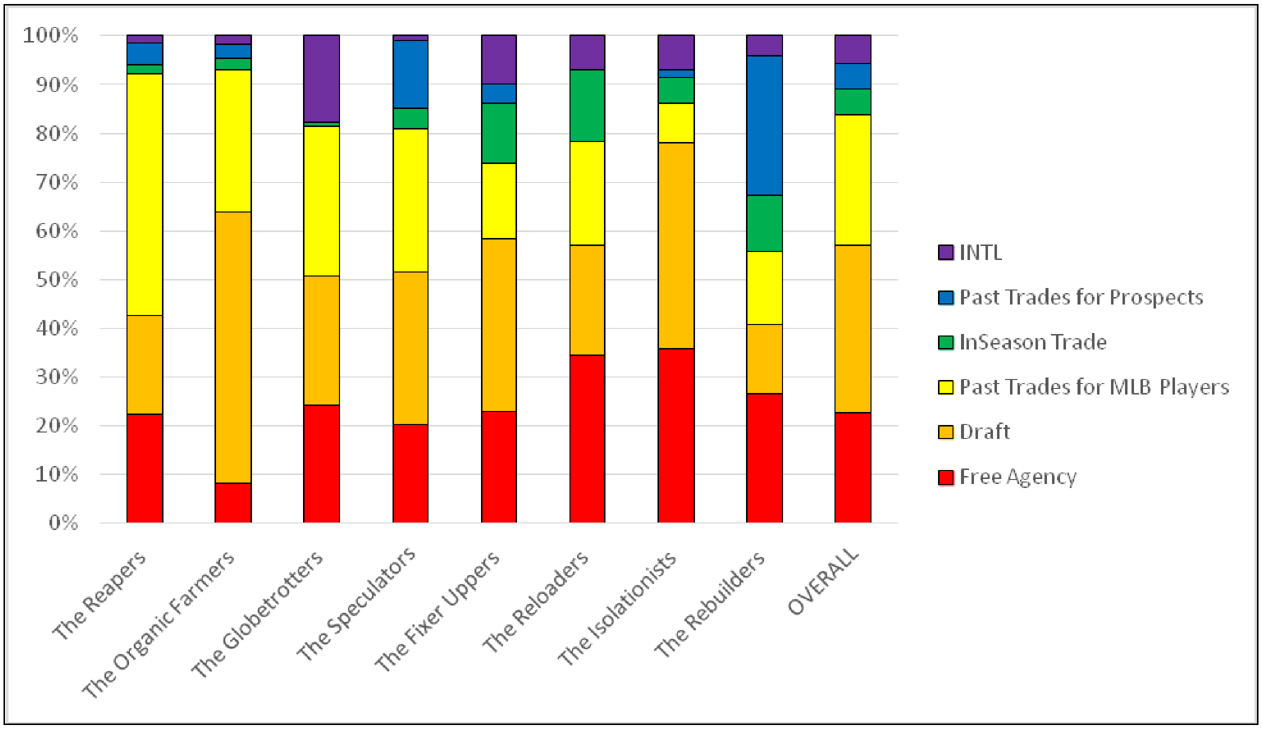
It appears that the clusters are fairly evenly sized, ranging from eight teams (the Organic Farmers) to two teams (the Rebuilders). Table 3 lists the teams in each cluster, as well as my description of the "defining feature" of each.

Table 3: The Clusters, Interpreted

Cluster	Defining Feature	Members
A. The Reapers	Willing to **trade for all stars**	77 NYY; 78 NYY; 80 PHI; 82 STL; 01 ARI; 04 BOS
B. The Organic Farmers	Heavily reliant on **draft**, and not on free agency	79 PIT; 83 BAL; 84 DET; 85 KCR; 87 MIN; 90 CIN; 02 ANA; 08 PHI
C. The Globetrotters	More **international** players (esp. pitchers) than average	81 LAD; 96 NYY; 98 NYY; 99 NYY; 07 BOS
D. The Speculators	Successfully **traded for prospects** a few years prior	86 NYM; 91 MIN; 92 TOR; 93 TOR; 05 CHW; 06 STL
E. The Fixer Uppers	Acquired playoff starter in **mid-season trade**	88 LAD; 11 STL; 15 KCR; 17 HOU
F. The Reloaders	High **free agency** and **mid-season trade** acquisitions; not reliant on draft	89 OAK; 97 FLA; 00 NYY
G. The Isolationists	Core of team built through **draft** and **free agency**, not trades	95 ATL; 09 NYY; 10 SFG; 12 SFG; 13 BOS; 14 SFG
H. The Rebuilders	Old team **traded away for prospects**, complemented with **free agents** and **mid-season trades**	03 FLA; 16 CHC

It is worth noting too that the defining feature of a cluster (for example, the prevalence of international players amongst the Globetrotters), does not necessarily signify that this is the primary acquisition type within that cluster. It just means that this is what is shared amongst the member teams. To get an idea of what the acquisition profile looks like for each cluster, see Table 4 below:

Table 4: WAR Totals by Acquisition Type

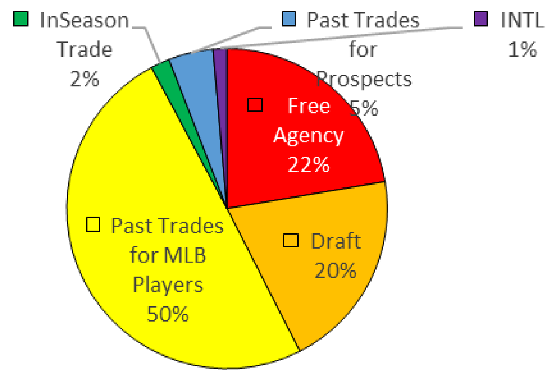


The table shows that the majority of players within each cluster were acquired via free agency, the draft, or past trades for MLB Players. As you can see though, clusters like the Globetrotters or Speculators are defined by their high relative prevalence (or absence) of secondary acquisition types.

4. The Individual Clusters

A. The Reapers

Defining Feature	Willing to **trade for all stars**
Member Teams	77 NYY; 78 NYY; 80 PHI; 82 STL; 01 ARI; 04 BOS
Team MVPs (Highest WAR)	77 NYY: Graig Nettles (5.5): Trade 78 NYY: Ron Guidry (9.6): Draft 80 PHI: Steve Carlton (10.2): Trade 82 STL: Lonnie Smith (6.1): Trade 01 ARI: Randy Johnson (10.0): FA 04 BOS: Curt Schilling (7.9): Trade

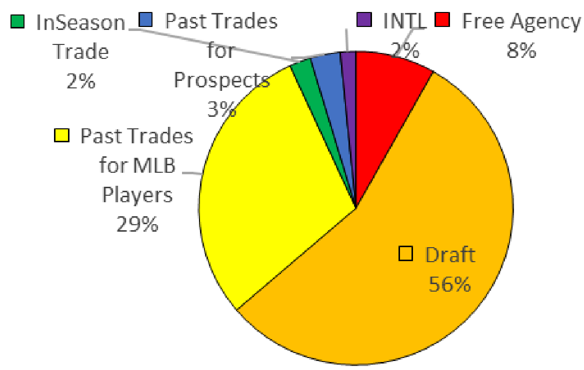


The Reapers, more than any other cluster, are defined by their reliance on trades to build their roster -- specifically, trades for established stars. In many cases, these teams traded away prospects or cash to obtain them. There are two notable subgroups within this cluster. The first comprises the 2001 Diamondbacks and 2004 Red Sox. While both teams relied heavily on trade acquisitions (46% of their total WAR), they also were stocked with high profile free-agent signings and had little to no representation from the draft. In fact, the only internally drafted player on either World Series roster was Trot Nixon, who contributed a negligible 0.9 WAR in 2004. If any team ever "purchased" a World Series title in the past 40 years, it was one of these two. Interestingly enough, the fact that Curt Schilling was a star for both may suggest that Theo Epstein, the general manager of the 2004 Red Sox, wanted to repeat the success the 2001 Diamondbacks had had against the once invincible Yankees.

The remaining teams form the second subgroup, and what is notable is that these are four of the six oldest teams considered in this analysis. More than 50% of the WAR from these four teams was acquired via trades, but instead of complementing this talent with free agents like the first subgroup, they were able to utilize the talent they had drafted and developed internally. This point will be illustrated in more detail later, but most World Series champions prior to 1988 did not rely heavily on free agency. Therefore, like the 2001 Diamondbacks and 2004 Red Sox, the 1977-78 Yankees, 1980 Phillies, and 1982 Cardinals can be viewed as some of the more aggressively constructed teams of their era.

B. The Organic Farmers

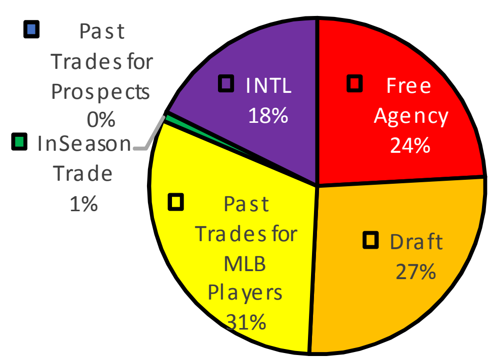
Defining Feature	Heavily reliant on **draft**, and not on free agency
Member Teams	79 PIT; 83 BAL; 84 DET; 85 KCR; 87 MIN; 90 CIN; 02 ANA; 08 PHI
Team MVPs (Highest WAR)	79 PIT: Dave Parker (6.7): Draft 83 BAL: Cal Ripken (8.2): Draft 84 DET: Alan Trammell (6.7): Draft 85 KCR: George Brett (8.3): Draft 87 MIN: Frank Viola (8.1): Draft 90 CIN: Jose Rijo (5.7): Trade 02 ANA: Darin Erstad (6.4): Draft 08 PHI: Chase Utley (9.0) : Draft



As mentioned, most World Series Champions built prior to 1988 were constructed primarily via trades or the draft. While the Reapers relied heavily on trades to complement (or supplement) their drafted players, the Organic Farmers' best players were almost always drafted and internally developed. This is the simplest and most distinct cluster; no other segment relied so much on the draft and so little on free agency (in fact, the 1990 Reds had no free agents on their World Series roster at all). While most of its constituent teams won prior to 1991, the 2002 Angels and 2008 Phillies proved that even in the age of rampant free agency, building through the draft was not an antiquated strategy for success. With stars like Chase Utley, Jimmy Rollins, Ryan Howard, and Cole Hamels, 57% of the Phillies WAR in 2008 was from drafted players. What is commendable is that the Phillies did not have to tank in order to acquire those four players; all were drafted after the Phillies had finished in at least third place. The Organic Farmers are also the largest cluster in the study with eight teams, which suggests that while not sexy, good scouting and patience can (and has) paid off many times.

C. The Globetrotters

Defining Feature	More **international** players (especially pitchers) than average
Member Teams	81 LAD; 96 NYY; 98 NYY; 99 NYY; 07 BOS
Team MVPs (Highest WAR))	81 LAD: Fernando Valenzuela (4.8): Int'l 96 NYY: Andy Pettitte (5.6): Draft 98 NYY: Derek Jeter (7.5): Draft 99 NYY: Derek Jeter (8.0): Draft 07 BOS: Josh Beckett (6.5): Trade



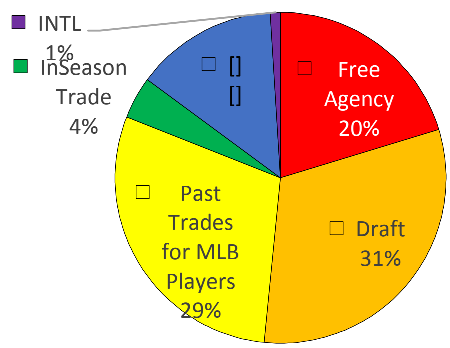
The Globetrotters are essentially the late 90's Yankees, and it is slightly surprising that the defining feature for a team famous for aggressive spending is actually its international scouting/acquisitions. But look closer and it is not quite so strange: Mariano Rivera, Orlando "El Duque" Hernandez, and -- most importantly -- Bernie Williams were all international signings. In fact, if not for Derek Jeter, Bernie Williams might have been the homegrown jewel of the Yankees' renaissance. From 1995 to 2002, he averaged 5.2 WAR per season (from

1996 to 2002, Jeter averaged 5.3 WAR), and he was 3rd, 4th, and 2nd in WAR on the Yankees in 1996, 1998, and 1999 respectively. The Yankees' international prowess was not unique to this era either; even into the 2000s, the Yankees continued to sign great international players, including Hideki Matsui, Alfonso Soriano and Robinson Cano.

A subtle point about this cluster is that it is actually defined by the presence of strong international *pitchers*, which is why the 1981 Dodgers (Fernando Valenzuela) and 2007 Red Sox (Daisuke Matsuzaka, Hideki Okajima) are included and the 2017 Astros (Jose Altuve, Yuli Gurriel) and 2009 Yankees (Cano, Matsui) are not.

D. The Speculators

Defining Feature	Successfully **traded for prospects** a few years prior
Member Teams	86 NYM; 91 MIN; 92 TOR; 93 TOR; 05 CHW; 06 STL
Team MVPs (Highest WAR)	86 NYM: Keith Hernandez (5.5): Trade 91 MIN: Kevin Tapani (6.8): Prospect trade 92 TOR: Roberto Alomar (6.6): Trade 93 TOR: John Olerud (7.7): Draft 05 CHW: Mark Buerhle (4.8): Draft 06 STL: Albert Pujols (8.5): Draft



While trading current players for future prospects has always been a viable strategy, it was not until the late 80's that these future stars actually started becoming key players on World Series teams. The 1991 World Series, famously known as the "worst to first" series, featured two of these players: Kevin Tapani of the Twins and John Smoltz of the Braves. Both had been acquired during the late 80's as minor leaguers via midseason trades that sent star pitchers (Frank Viola and Doyle Alexander respectively) to better teams. Why trading for prospects only started paying off during this time is beyond the scope of this paper, but the strategy does appear to be an evolution of the Reapers' and Organic Farmers' strategies. The 1992-93 Blue Jays and 1991 Twins still relied heavily on the draft and trades (for major leaguers, not prospects), but what separated them from previous winners was the presence of strategic free agents (Dave Winfield and Paul Molitor for the Blue Jays, Chili Davis for the Twins, Jack Morris for both) and prospects acquired via trade (Juan Guzman for the Jays and Tapani for the Twins). There are actually a lot of similarities between this cluster and the Rebuilders. However, while the Speculators relied mostly on the draft and other trades, The Rebuilders' success was much more dependent on free agency.

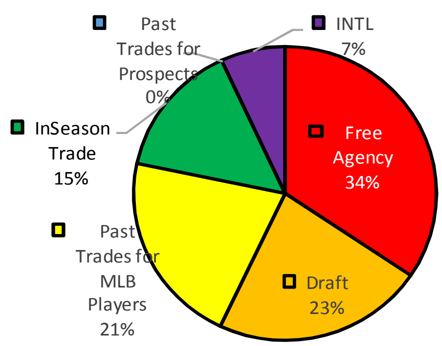
E. The Fixer-Uppers

Defining Feature	Acquired playoff starter in **mid-season trade**
Member Teams	88 LAD; 11 STL; 15 KCR; 17 HOU
Team MVPs (Highest WAR)	88 LAD: Orel Hershiser (7.2): Draft 11 STL: Albert Pujols (5.3): Draft 15 KCR: Lorenzo Cain (7.2): Prospect trade 17 HOU: Jose Altuve (8.3): International

At first glance, it appears the reason these four teams are grouped together is because they all traded for a starting pitcher mid-season to bolster their rotation. In fact, John Tudor (1988 Dodgers), Edwin Jackson (2011 Cardinals), Johnny Cueto (2015 Royals), and Justin Verlander (2017 Astros) all started at least one game in the World Series for their respective teams. But looking deeper, there is a common thread that is not initially obvious, especially between the 2015 Royals and the 2017 Astros (and it's not that Salvador Perez and Jose Altuve played Little League together in Venezuela). It's that these two teams did not adhere to one particular building strategy, apparently having no preference for how they obtained players. In other words, they'd both acquired significant talent via all six acquisition channels. Winning in this fashion requires flexibility and self-awareness from a front office. General Managers need to know when to give up on players who are not performing, no matter what has been invested in them. They also need to know when their team has its best chance to win, and "strike while the iron is hot". Operating in this way requires spontaneity and improvisation, and may be frightening for those who need planned structure. However, great results can be achieved (cheaply!) if managed correctly, as demonstrated by the Astros' and Royals' adroit maneuvering.

F. The Reloaders

Defining Feature	High **free agency** and **mid-season trade** acquisitions; not reliant on draft
Member Teams	89 OAK; 97 FLA; 00 NYY
Team MVPs (Highest WAR)	89 OAK: Rickey Henderson (8.7): In-season trade 97 FLA: Kevin Brown (7.0): FA 00 NYY: Jorge Posada (5.5): Draft

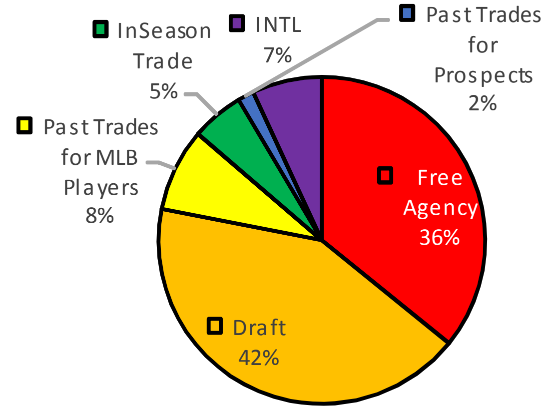


As the name implies, these three seasons featured already-great teams that improved themselves through recent free agent signings and mid-season trades. In fact two of the teams (The 1989 Oakland Athletics and 2000 New York Yankees) had played in the World Series the season before, but due to unforeseen complications needed to bolster their lineups to maintain their dominance (acquiring Rickey Henderson and Dave Justice, respectively). In fact, the trade for Justice is why the 2000 version of the Yankees is not a member of the Globetrotters with the other late 90's Yankees teams. Those teams never struggled during their regular seasons and did not need to make in-season improvements.

Like the A's and Yankees, the 1997 Marlins were heavily comprised of free agents, and had also made in-season trade acquisitions (Darren Daulton and Craig Counsell). However, many of their best players had been acquired as free agents only a short time before (Moises Alou, Alex Fernandez, Kevin Brown), with only one significant homegrown player on the roster (Charles Johnson). Given that they were just a five-year-old expansion team, free agency may have been the only avenue they had to win a World Series so quickly. Unfortunately though, the dust hadn't even settled on their championship trophy before the entire team was dismantled and traded away. The quick rise and fall of the 1997 Marlins, and subsequent success of the 2003 Marlins, introduced and briefly (albeit speciously) validated a strategy that violated the ethics of competitive sportsmanship: tanking. However, three (more) roster purges, fourteen seasons, and zero playoff appearances later, the Marlins have been unequivocally branded as the epitome of operational dysfunction and malpractice.

G. The Isolationists

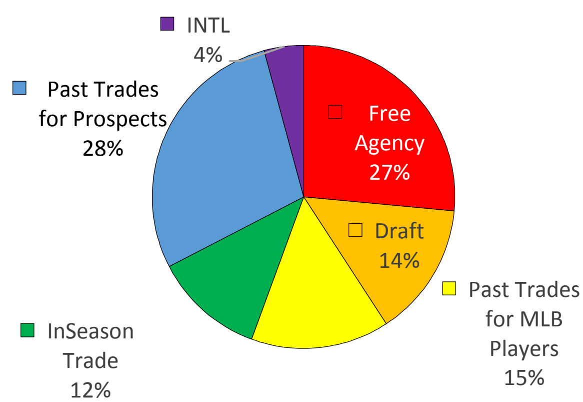
Defining Feature	Core of team built through **draft** and **free agency**, not trades
Member Teams	95 ATL; 09 NYY; 10 SFG; 12 SFG; 13 BOS; 14 SFG
Team MVPs (Highest WAR)	95 ATL: Greg Maddux (9.7): FA 09 NYY: Derek Jeter (6.5): Draft 10 SFG: Aubrey Huff (5.7): FA 12 SFG: Buster Posey (7.3): Draft 13 BOS: Dustin Pedroia (6.3): Draft 14 SFG: Madison Bumgarner (5.3): Draft 14 SFG: Buster Posey: (5.3): Draft



The Isolationists are teams built almost exclusively through free agency and the draft. By definition, trades are transactions that should benefit both teams involved. The Isolationists are self-reliant though, and did not have to help others to help themselves. While on average, 36 percent of the WAR from teams in this cluster comes from free agency and 42 percent comes from the draft, there is some variability between these numbers amongst the member teams. The 1995 Braves and 2012 Giants teams were built primarily via the draft (if not for free agent Greg Maddux, the 1995 Braves team would have been a member of the Organic Gardener cluster), while both the 2009 Yankees and 2013 Red Sox were very heavy on free agents. Winning this way is very simple and straightforward, but requires either great scouting/development or a lot of money (or both).

H. The Rebuilders

Defining Feature	Old team **traded away for prospects**, complemented with **free agents** and **mid-season trades**
Member Teams	03 FLA; 16 CHC
Team MVPs (Highest WAR)	03 FLA: Ivan Rodriguez (4.5): FA 16 CHC: Kris Bryant (7.7): Draft



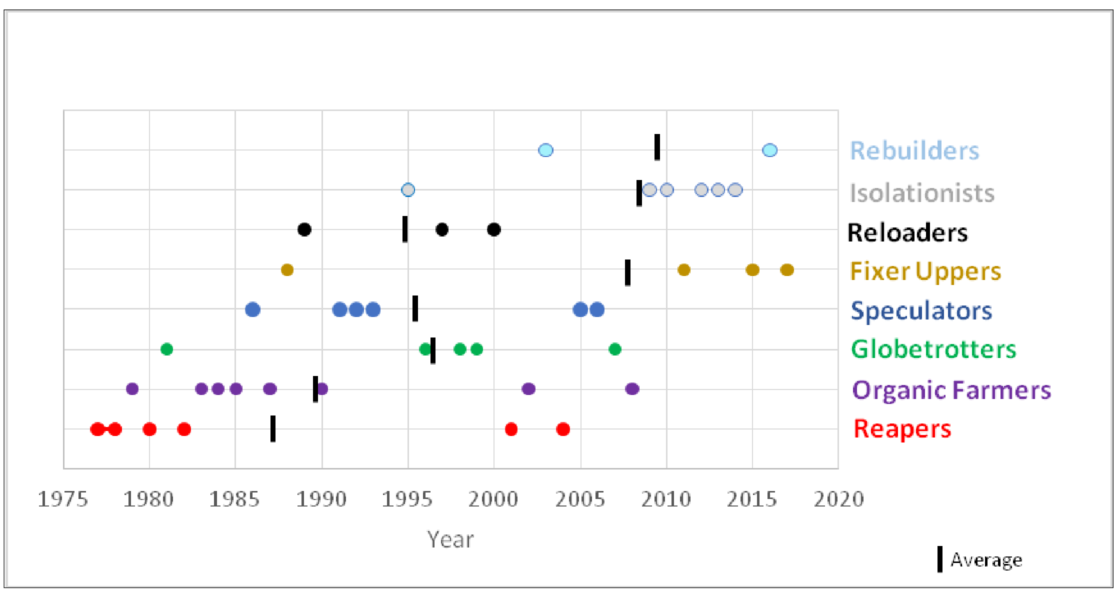
Tanking may be a tenable strategy in theory, but in practice it has never produced a World Series champion. One could make the argument that the 2003 Marlins were the result of the teardown of the 1997 team, but rigorous inspection provides little supporting evidence that this was true. Of the trades made after the 1997 season (Kevin Brown, Gary Sheffield, Charles Johnson, Bobby Bonilla, and Moises Alou were all traded away), only one yielded a player who contributed significantly to the 2003 championship team (Kevin Brown for Derrek Lee). Outside of Josh Beckett (the 2nd overall pick in the 1999 draft), the Marlins tanking produced no other players on this 2003 team. Most other young prospects were acquired in shrewd minor league swaps (Brad Penny, Dontrelle Willis, and Mike Lowell), and the team's best player according to WAR (Ivan Rodriguez) was a free agent signing. This suggests that the Marlins did not have to blow up their 1997 team in order to produce this 2003 team. However, given the team had been blown up, Marlins management did do a great job of rebuilding the roster.

The 2003 Marlins would be the anomaly in this group of forty teams if not for the 2016 Cubs, who also had also traded for many of their best players as minor league prospects (Anthony Rizzo, Addison Russell, and Kyle Hendricks). While both of these teams are similar to the Speculators, where they differ is in their reliance on free agency and in-season acquisitions instead of the draft or trades (especially the Cubs). Unlike the Marlins, the Cubs had never really tanked by design; it was more due to negligence, as Chicago was already at the bottom of the standings when they traded for Russell and Hendricks. However, they did defer their urgency to start winning again. They could have started signing players immediately when Theo Epstein became GM in 2011, but did not, deciding to wait for their prospects to blossom before aggressively spending on the free agent market. However, when they were ready, they built very quickly. With the exception of Jon Lester (who was acquired during the 2014 offseason), all significant free agents on the 2016 Cubs had been acquired or re-signed the previous offseason (John Lackey, Ben Zobrist, Jason Heyward, and Dexter Fowler).

5. Discussion

Not only did this analysis shed light on the different ways championship teams have been constructed over the past forty years, but it also showed how strategies have changed in that interval. Table 5 presents the timeline:

Table 5: Championship Strategy Timeline



There appears to be a gradual evolution of roster building, starting with the Reapers, followed by the Organic Farmers and the Speculators, before the Yankees reigned supreme in the late 90's. There was no clear pattern again until 2009 when the Yankees defined the new normal with the Isolationists. However, by 2015 the rules had changed again with the analytically driven Fixer-Uppers.

The 2017 Astros, arguably the definitive team of the Fixer-Uppers, have become the new face of analytics in baseball (Arguably, the "Moneyball" A's lost whatever claim they had to throne after imploding following the trade of Yoenis Cespedes to the Red Sox in 2014). The Astros have also, partly thanks to Sports Illustrated,[6] become the face of tanking, which has turned them into villains, harbingers of an unwanted future. The problem with tanking is that in any given season, it's possible that up to 17 to 20 percent of teams are admittedly not trying to be competitive (as was the case in 2017, when the Braves, White Sox, Phillies, Tigers, and Padres all fielded AAA caliber lineups), opting instead to cut costs and use their losing records to hoard high draft picks. Competitive integrity aside, fans of tanking teams must endure the pain of waiting at least three to four seasons to see their team succeed again. But what if they realized that the tanking was unnecessary, as this analysis suggests, or worse did not pay off in a World Series championship?

The analysis presented here finds no evidence that any World Series winner of the past forty years succeeded because they had purposely tanked in the few seasons prior. The Rebuilders' teams (the 2003 Marlins and 2016 Cubs) were the most likely candidates (and not the 2017 Astros), but closer inspection revealed that their recent bad seasons were either unnecessary (in the case of the Marlins) or the result of many years of institutional ineptitude (the Cubs). Neither team benefited from a deliberate teardown and rebuild (although the Cubs may have benefitted from purposefully not improving an already-bad team).

The tanking of the 2011-2013 Astros appears to be a red herring. Their three 50-win seasons yielded only one player who contributed significantly to their 2017 title: Carlos Correa. The other keystones of the team -- Jose Altuve, George Springer, Dallas Keuchel, and Justin Verlander -- were all acquired before the "intentional" tanking began, or after the team had matured into a winner. As stated in the description of the Fixer-Uppers, The Astros won the title in 2017 because they were flexible, made careful decisions and took smart risks (and traded for Justin Verlander). Finishing in last place from 2011 to 2013 had less of an impact in real life than it has in the popular narrative. However, having the smart front office that they do, the Astros would be wise not to correct this public misperception. They have now inspired copycats, and the fewer competitive teams there are in the league, the greater their odds are to repeat as World Series champions.

Mark Kindem, mark.kindem@gmail.com ♦

[6] https://www.si.com/mlb/2017/10/24/houston-astros-jeff-luhnow-jim-crane-tanking

Back issues

Back issues of "By the Numbers" are available at the SABR website, at http://sabr.org/research/statistical-analysis-research-committee-newsletters, and at editor Phil Birnbaum's website, www.philbirnbaum.com .

The SABR website also features back issues of "Baseball Analyst", the sabermetric publication produced by Bill James from 1981 to 1989. Those issues can be found at http://sabr.org/research/baseball-analyst-archives.

Submissions

Phil Birnbaum, Editor

Submissions to *By the Numbers* are, of course, encouraged. Articles should be concise (though not necessarily short), and pertain to statistical analysis of baseball. Letters to the Editor, original research, opinions, summaries of existing research, criticism, and reviews of other work are all welcome.

Articles should be submitted in electronic form, preferably by e-mail. I can read most word processor formats. If you send charts, please send them in word processor form rather than in spreadsheet. Unless you specify otherwise, I may send your work to others for comment (i.e., informal peer review).

I usually edit for spelling and grammar. If you can (and I understand it isn't always possible), try to format your article roughly the same way BTN does.

I will acknowledge all articles upon receipt, and will try, within a reasonable time, to let you know if your submission is accepted.

Send submissions to Phil Birnbaum, at philbirnbaum@outlook.com.

"By the Numbers" mailing list

SABR members who have joined the Statistical Analysis Committee will receive e-mail notification of new issues of BTN, as well as other news concerning this publication.

The easiest way to join the committee is to visit http://members.sabr.org, click on "my SABR," then "committees and regionals," then "add new" committee. Add the Statistical Analysis Committee, and you're done. You will be informed when new issues are available for downloading from the internet.

If you would like more information, send an e-mail to Phil Birnbaum, at philbirnbaum@outlook.com. If you don't have internet access, we will send you BTN by mail; write to Phil at 88 Westpointe Cres., Ottawa, ON, K2G 5Y8.