Playoff Teams - Classification Estimation

Connor Gooding

8/22/2017

Synopsis

This analysis compares every single-season team performance in MLB history since 1903, the year of the inaugural World Series, identifies different tiers of postseason success, and explores the characteristics of these playoff success clusters.

Data Acquisition

All data for single-season team performances in this report come from the FanGraphs batting and pitching leaderboards. The raw .csv files are hidden behind Javascript functions, and so direct links could not be used to donwload the files directly into RStudio.

It is also important to note that the 'Export Data' function of these leaderboards names all of the exported .csv files the same name (FanGraphs Leaderboard.csv). To better identify the different tables being pulled into this report, some considerable renaming was done to the downloaded .csv files.

```
batting_teams_dash <- read.csv('batting_teams_dash.csv')
pitching_teams_dash <- read.csv('pitching_teams_dash.csv')
ws_winners <- read.csv("ws_winners.csv", stringsAsFactors = F)
ws_losers <- read.csv("ws_losers.csv", stringsAsFactors = F)
```

```
str(batting_teams_dash)
```

```
## 'data.frame':    2865 obs. of  21 variables:
##  $ Season: int  1927 1976 1931 1939 1969 2001 1975 1884 1948 1973 ...
##  $ Team  : Factor w/ 118 levels "Alleghenys","Americans",..: 118 91 118 118 72 59 91 61 53 72 ...
##  $ G     : int  1725 2135 1751 1550 1981 2043 2148 1023 1893 1707 ...
##  $ PA    : int  6222 6538 6473 6129 6328 6474 6419 4466 6201 6335 ...
##  $ HR    : int  158 141 155 166 175 169 124 32 155 119 ...
##  $ R     : int  976 857 1067 967 779 927 840 887 840 754 ...
##  $ RBI   : int  908 802 990 904 722 881 779 NA 801 692 ...
##  $ SB    : int  90 210 139 72 82 174 168 NA 54 146 ...
##  $ BB.   : Factor w/ 122 levels "0.0 %","0.2 %",..: 21 22 34 32 18 118 26 64 22 20 ...
##  $ K.    : Factor w/ 229 levels "","0.7 %","0.9 %",..: 228 52 216 219 41 67 57 1 223 33 ...
##  $ ISO   : num  0.181 0.144 0.16 0.164 0.148 0.157 0.13 0.102 0.15 0.123 ...
##  $ BABIP : num  0.324 0.309 0.309 0.295 0.281 0.32 0.303 NA 0.292 0.287 ...
##  $ AVG   : num  0.307 0.28 0.297 0.287 0.265 0.288 0.271 0.292 0.282 0.266 ...
##  $ OBP   : num  0.384 0.357 0.383 0.374 0.343 0.36 0.353 0.321 0.36 0.345 ...
##  $ SLG   : num  0.488 0.424 0.457 0.451 0.414 0.445 0.401 0.394 0.431 0.389 ...
##  $ wOBA  : num  0.4 0.357 0.387 0.38 0.341 0.347 0.345 0.326 0.369 0.334 ...
##  $ wRC.  : int  126 120 124 113 112 116 111 132 114 111 ...
##  $ BsR   : num  -14.2 19 -3.4 -4.6 1.7 16.2 20.8 0 -10.4 7.3 ...
##  $ Off   : num  203.7 166.6 206.4 104.6 86.9 ...
##  $ Def   : num  76.2 39.2 18.8 129.7 150.9 ...
##  $ WAR   : num  49.3 45.9 45 44.7 44.6 44.4 43.2 43 42 41.2 ...
```

```
str(pitching_teams_dash)
```

```
## 'data.frame':    2865 obs. of  19 variables:
##  $ Season: int  1996 2003 1997 2002 2011 1999 1970 2003 1998 1971 ...
##  $ Team  : Factor w/ 118 levels "Alleghenys","Americans",..: 15 118 15 118 79 15 34 38 15 114 ...
##  $ W     : int  96 101 101 103 102 103 84 85 106 79 ...
##  $ L     : int  66 61 61 58 60 59 78 77 56 83 ...
##  $ SV    : int  46 49 37 53 47 45 25 58 45 32 ...
##  $ G     : int  570 530 536 495 556 556 367 600 516 421 ...
##  $ GS    : int  162 163 162 161 162 162 162 162 162 162 ...
##  $ IP    : num  1469 1462 1465 1452 1477 ...
##  $ K.9   : num  7.63 6.89 7.34 7.04 7.92 7.32 6.27 7.96 7.71 6.06 ...
##  $ BB.9  : num  2.76 2.31 2.76 2.5 2.46 3.1 2.98 3.25 2.92 2.9 ...
##  $ HR.9  : num  0.74 0.89 0.68 0.89 0.73 0.87 0.9 0.78 0.73 0.62 ...
##  $ BABIP : num  0.291 0.302 0.283 0.293 0.286 0.289 0.283 0.28 0.285 0.276 ...
##  $ LOB.  : Factor w/ 361 levels "-5.8 %","14.2 %",..: 292 283 320 286 352 312 298 349 323 311 ...
##  $ GB.   : Factor w/ 100 levels "","38.2 %","38.6 %",..: 1 32 1 33 70 1 1 85 1 1 ...
##  $ HR.FB : Factor w/ 80 levels "","10.0 %","10.1 %",..: 1 74 1 71 64 1 1 7 1 1 ...
##  $ ERA   : num  3.54 4.03 3.18 3.89 3.02 3.65 3.76 3.16 3.25 3.13 ...
##  $ FIP   : num  3.5 3.66 3.45 3.62 3.24 3.85 3.59 3.56 3.53 3.03 ...
##  $ xFIP  : num  NA 3.92 NA 3.86 3.41 NA NA 3.63 NA NA ...
##  $ WAR   : num  29.5 28.6 28.5 28.2 28.1 27.8 27.1 27 27 26.9 ...
```

```
batting_teams_dash <- batting_teams_dash %>% filter(Season >= 1903)
pitching_teams_dash <- pitching_teams_dash %>% filter(Season >= 1903)

batting_teams_dash$SeasonTeam <- with(batting_teams_dash, paste(Season, Team))
pitching_teams_dash$SeasonTeam <- with(pitching_teams_dash, paste(Season, Team))
```

```
playoff_results <- batting_teams_dash %>% arrange(Season, Team) %>% select(SeasonTeam)
playoff_results$WonWS <- rep(0, length(playoff_results$SeasonTeam))
playoff_results[playoff_results$SeasonTeam %in% ws_winners$SeasonTeam, 2] <- 1

batting_teams_dash$WonWS <- batting_teams_dash$SeasonTeam %in% ws_winners$SeasonTeam
pitching_teams_dash$WonWS <- pitching_teams_dash$SeasonTeam %in% ws_winners$SeasonTeam

batting_teams_dash$LostWS <- batting_teams_dash$SeasonTeam %in% ws_losers$SeasonTeam
pitching_teams_dash$LostWS <- pitching_teams_dash$SeasonTeam %in% ws_losers$SeasonTeam


batting_teams_dash <- batting_teams_dash %>% select(1:2, SeasonTeam, WonWS, LostWS, G:WAR)
pitching_teams_dash <- pitching_teams_dash %>% select(1:2, SeasonTeam, WonWS, LostWS, W:WAR)
```