post02-June-Namgung

June Namgung
12/2/2017

Analyzing NBA Data Based on Efficiency

Introduction

We have been dealing with NBA data for the entire semester but since I am a die hard basketball fan, I need some more. From homework 3 that we did before, we have the rank of team efficiency. Many people say team should be efficient enough to win and get a shape. My curiosity started from here. Does team efficiency mean that a lot? I suppose team standings will follow each team's efficiency. So, I tried to gather 3 latest NBA regular seasons and playoffs' team stats to find whether efficiency held a role a lot.

Method

I am focusing on basketball efficiency itself. There are many basketball stats related to efficiency but I picked two, offensive efficiency rating and defensive efficiency rating. And then I will take a difference between offensive and defensive efficiency ratings in order to get pure margins from those. As I mentioned before, I suppose team standings will follow the efficiency ratings.

First, let's import some libraries we need.

```r
#importing libraries
library(ggplot2)
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.4.2

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

And then we import the datasets, 14-15, 15-16, 16-17 NBA regular season and 15,16,17 NBA playoffs' team stats from nba official site.

```
#importing the datasets
reg14 <- read.csv('../data/1415REG.csv', stringsAsFactors = FALSE)
reg15 <- read.csv('../data/1516REG.csv', stringsAsFactors = FALSE)
reg16 <- read.csv('../data/1617REG.csv', stringsAsFactors = FALSE)
po15 <- read.csv('../data/15PO.csv', stringsAsFactors = FALSE)
po16 <- read.csv('../data/16PO.csv', stringsAsFactors = FALSE)
po17 <- read.csv('../data/17PO.csv', stringsAsFactors = FALSE)
reg14
```

```
##    Rk                   Team  G    MP   FG  FGA   FG. X3P X3PA  X3P.  X2P
## 1   1  Golden State Warriors 82 19730 3410 7137 0.478 883 2217 0.398 2527
## 2   2   Los Angeles Clippers 82 19730 3228 6830 0.473 827 2202 0.376 2401
## 3   3       Dallas Mavericks 82 19880 3255 7036 0.463 732 2082 0.352 2523
## 4   4        Toronto Raptors 82 19855 3108 6829 0.455 726 2060 0.352 2382
## 5   5  Oklahoma City Thunder 82 19830 3184 7119 0.447 632 1864 0.339 2552
## 6   6        Houston Rockets 82 19805 3032 6832 0.444 933 2680 0.348 2099
## 7   7      San Antonio Spurs 82 19955 3208 6854 0.468 677 1847 0.367 2531
## 8   8    Cleveland Cavaliers 82 19780 3089 6739 0.458 826 2253 0.367 2263
## 9   9 Portland Trail Blazers 82 19855 3175 7049 0.450 807 2231 0.362 2368
## 10 10          Atlanta Hawks 82 19730 3121 6699 0.466 818 2152 0.380 2303
## 11 11           Phoenix Suns 82 19880 3178 7038 0.452 698 2048 0.341 2480
## 12 12         Denver Nuggets 82 19880 3099 7158 0.433 660 2032 0.325 2439
## 13 13         Boston Celtics 82 19880 3193 7211 0.443 660 2021 0.327 2533
## 14 14       Sacramento Kings 82 19855 3010 6617 0.455 461 1350 0.341 2549
## 15 15          Chicago Bulls 82 19880 3001 6797 0.442 645 1825 0.353 2356
## 16 16   New Orleans Pelicans 82 19780 3108 6795 0.457 586 1583 0.370 2522
## 17 17     Washington Wizards 82 19955 3139 6790 0.462 497 1381 0.360 2642
## 18 18        Detroit Pistons 82 19830 3041 7038 0.432 703 2043 0.344 2338
## 19 19     Los Angeles Lakers 82 19930 3054 7020 0.435 532 1546 0.344 2522
## 20 20      Memphis Grizzlies 82 19905 3097 6763 0.458 423 1246 0.339 2674
## 21 21          Brooklyn Nets 82 19930 3069 6804 0.451 541 1633 0.331 2528
## 22 22        Milwaukee Bucks 82 19930 3083 6722 0.459 545 1500 0.363 2538
## 23 23 Minnesota Timberwolves 82 19805 2986 6820 0.438 406 1223 0.332 2580
## 24 24         Indiana Pacers 82 19855 2998 6824 0.439 612 1740 0.352 2386
## 25 25          Orlando Magic 82 19755 3076 6792 0.453 554 1598 0.347 2522
## 26 26              Utah Jazz 82 19705 2900 6492 0.447 610 1781 0.343 2290
## 27 27             Miami Heat 82 19730 2885 6330 0.456 556 1659 0.335 2329
## 28 28      Charlotte Hornets 82 19905 2913 6932 0.420 498 1566 0.318 2415
## 29 29     Philadelphia 76ers 82 19805 2765 6777 0.408 692 2160 0.320 2073
## 30 30        New York Knicks 82 19855 2882 6726 0.428 560 1614 0.347 2322
##    X2PA  X2P.   FT  FTA   FT.  ORB  DRB  TRB  AST STL BLK  TOV   PF  PTS
## 1  4920 0.514 1313 1709 0.768  853 2814 3667 2248 762 496 1185 1628 9016
## 2  4628 0.519 1468 2067 0.710  784 2711 3495 2031 640 409 1012 1749 8751
## 3  4954 0.509 1386 1843 0.752  858 2608 3466 1846 663 371 1062 1644 8628
## 4  4769 0.499 1585 2014 0.787  881 2526 3407 1701 615 357 1057 1712 8527
## 5  5255 0.486 1524 2020 0.754 1052 2844 3896 1681 598 454 1205 1829 8524
## 6  4152 0.506 1525 2133 0.715  958 2624 3582 1820 777 407 1366 1803 8522
## 7  5007 0.505 1368 1754 0.780  806 2772 3578 2000 657 444 1146 1564 8461
## 8  4486 0.504 1453 1934 0.751  911 2612 3523 1814 603 340 1171 1510 8457
## 9  4818 0.491 1272 1589 0.801  879 2881 3760 1799 525 372 1117 1494 8429
## 10 4547 0.506 1349 1735 0.778  715 2611 3326 2111 744 380 1167 1457 8409
## 11 4990 0.497 1343 1767 0.760  896 2643 3539 1659 700 385 1238 1744 8397
## 12 5126 0.476 1462 1991 0.734 1012 2653 3665 1788 641 367 1166 1882 8320
## 13 5190 0.488 1266 1678 0.754  910 2685 3595 2009 674 294 1133 1738 8312
## 14 5267 0.484 1829 2400 0.762  895 2728 3623 1667 550 324 1333 1696 8310
## 15 4972 0.474 1618 2067 0.783  959 2792 3751 1781 514 476 1145 1495 8265
## 16 5212 0.484 1345 1790 0.751  942 2621 3563 1806 553 510 1087 1530 8147
## 17 5409 0.488 1305 1758 0.742  862 2801 3663 1969 601 378 1233 1707 8080
## 18 4995 0.468 1292 1838 0.703 1051 2630 3681 1771 623 383 1099 1559 8077
## 19 5474 0.461 1433 1935 0.741  952 2647 3599 1715 578 366 1086 1741 8073
## 20 5517 0.485 1445 1869 0.773  856 2634 3490 1777 700 347 1094 1567 8062
## 21 5171 0.489 1359 1817 0.748  846 2627 3473 1716 576 340 1133 1579 8038
## 22 5222 0.486 1312 1734 0.757  876 2574 3450 1932 789 403 1373 1814 8023
## 23 5597 0.461 1638 2110 0.776  949 2406 3355 1771 668 327 1231 1571 8016
## 24 5084 0.469 1373 1817 0.756  856 2822 3678 1757 505 375 1147 1742 7981
## 25 5194 0.486 1141 1565 0.729  822 2607 3429 1692 647 314 1221 1714 7847
## 26 4711 0.486 1391 1929 0.721  988 2617 3605 1632 623 489 1256 1583 7801
## 27 4671 0.499 1438 1940 0.741  747 2461 3208 1626 642 372 1214 1636 7764
## 28 5366 0.450 1397 1867 0.748  820 2793 3613 1654 499 448  976 1494 7721
## 29 4617 0.449 1320 1953 0.676  978 2536 3514 1683 789 487 1453 1778 7542
## 30 5112 0.454 1211 1575 0.769  867 2443 3310 1746 575 382 1206 1768 7535
```

We need opponents team stats as well, to get defensive ratings.

```
reg14_opp <- read.csv('../data/opponent/1415REG_OPP.csv', stringsAsFactors = FALSE)
reg15_opp <- read.csv('../data/opponent/1516REG_OPP.csv', stringsAsFactors = FALSE)
reg16_opp <- read.csv('../data/opponent/1516REG_OPP.csv', stringsAsFactors = FALSE)
po15_opp <- read.csv('../data/opponent/15PO_OPP.csv', stringsAsFactors = FALSE)
po16_opp <- read.csv('../data/opponent/16PO_OPP.csv', stringsAsFactors = FALSE)
po17_opp <- read.csv('../data/opponent/17PO_OPP.csv', stringsAsFactors = FALSE)
```

These are the formulas that we will use.

OffRating = PTS * (100 / (FGA - OREB + TOV + (FTA * 0.44))))

DefRating = Opponent Points * (100 / (Opponent Field Goals Attempted - Opponent Off Rebounds + Opponent Turnovers + (Opponent Free Throws Attempted * 0.44))))

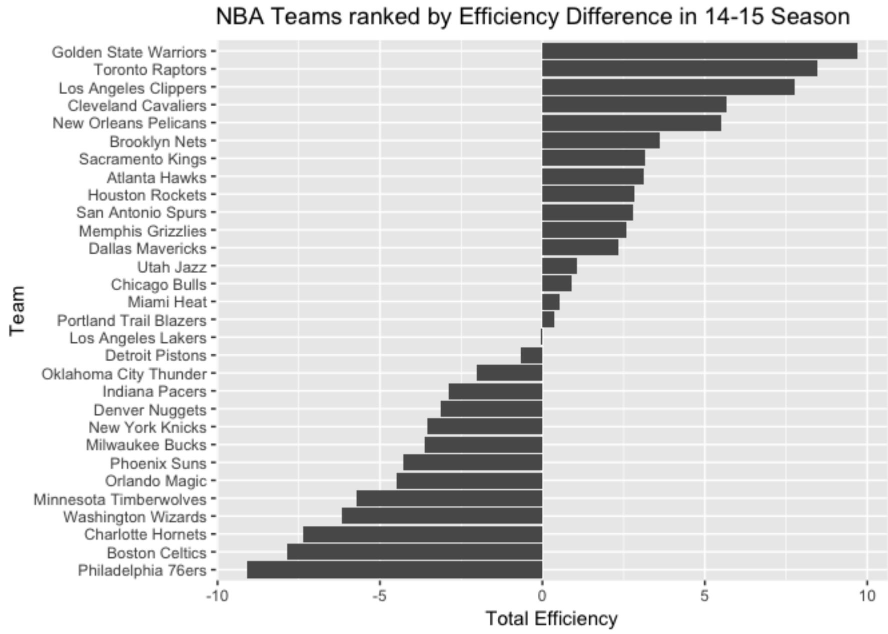
And I will call the difference between OffRating and DefRating as pure efficiency difference.

Let's calculate these stats right away from 14-15 NBA regular season.

```
reg14_rating <- reg14 %>%
  mutate(OffRating = PTS * (100 / (FGA - ORB + TOV + (FTA * 0.44))),
         DefRating = reg14_opp$PTS * (100 / (reg14_opp$FGA - reg14_opp$OREB + reg14_opp$TOV +(reg14_opp$FTA
reg14_rating <- reg14_rating[c(2,26,27)] #taking team names, offratings, defratings
reg14_rating <- reg14_rating %>%
  mutate(diff = OffRating - DefRating)
#sort pure efficiency difference by descending order
reg14_rating <- reg14_rating[order(reg14_rating$diff, decreasing = TRUE),]
reg14_rating
```

```
##                      Team OffRating DefRating        diff
## 1   Golden State Warriors 109.67089  99.98155  9.68933818
## 4         Toronto Raptors 108.05762  99.57730  8.48032354
## 2    Los Angeles Clippers 109.83398 102.05339  7.78058163
## 8     Cleveland Cavaliers 107.73303 102.08259  5.65043961
## 16   New Orleans Pelicans 105.42730  99.90466  5.52264212
## 21          Brooklyn Nets 101.86959  98.23824  3.63135623
## 14       Sacramento Kings 102.45346  99.29247  3.16098550
## 10          Atlanta Hawks 106.24937 103.13674  3.11262709
## 6         Houston Rockets 104.19978 101.35796  2.84182097
## 7       San Antonio Spurs 106.21711 103.42849  2.78862152
## 20      Memphis Grizzlies 103.05035 100.46232  2.58803001
## 3        Dallas Mavericks 107.16788 104.83814  2.32973895
## 26              Utah Jazz 102.52656 101.45342  1.07314226
## 15          Chicago Bulls 104.71994 103.83744  0.88249354
## 27             Miami Heat 101.48224 100.97266  0.50957939
## 9  Portland Trail Blazers 105.54509 105.18168  0.36341557
## 19     Los Angeles Lakers 100.84443 100.89459 -0.05015683
## 18        Detroit Pistons 102.30888 102.96958 -0.66069196
## 5   Oklahoma City Thunder 104.45054 106.47555 -2.02500883
## 24         Indiana Pacers 100.84048 103.71136 -2.87087377
## 12         Denver Nuggets 101.61162 104.72385 -3.11222735
## 30        New York Knicks  97.12555 100.67708 -3.55153209
## 22        Milwaukee Bucks 100.51416 104.15443 -3.64027355
## 11           Phoenix Suns 102.93620 107.20726 -4.27106218
## 25          Orlando Magic  99.58627 104.06804 -4.48177088
## 23 Minnesota Timberwolves  99.82068 105.54610 -5.72542254
## 17     Washington Wizards 101.83351 108.02060 -6.18709097
## 28      Charlotte Hornets  97.61704 104.97176 -7.35472210
## 13         Boston Celtics 101.70918 109.58564 -7.87645493
## 29     Philadelphia 76ers  92.98117 102.09178 -9.11061344
```

```
ggplot(data = reg14_rating, aes(x = reorder(Team, diff), y = diff)) +
  geom_bar(stat = 'identity') +
  labs(x = "Team", y = "Total Efficiency") +
  ggtitle("NBA Teams ranked by Efficiency Difference in 14-15 Season") +
  coord_flip()
```



(Visualizing efficiency difference.)

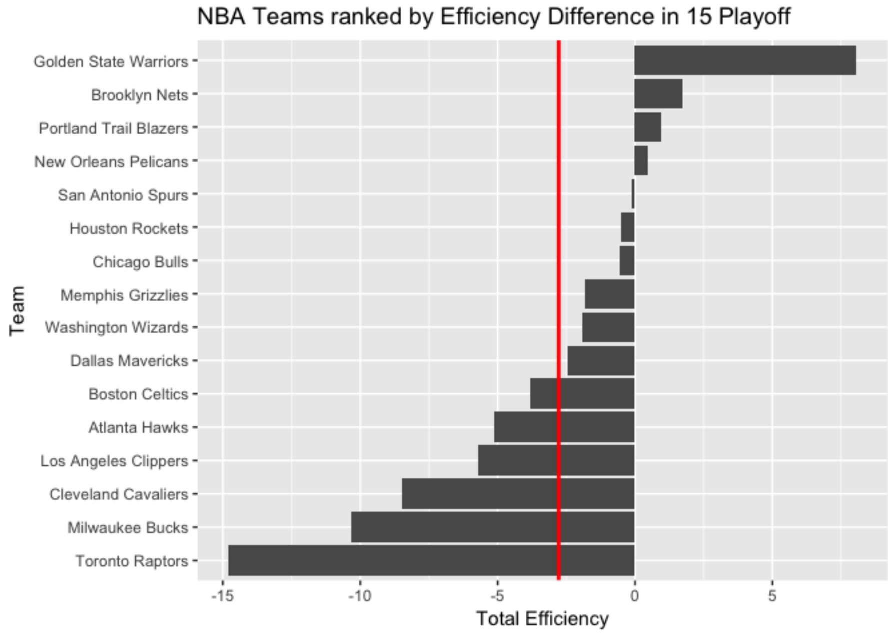
By looking at the table and the plot, pure efficiency difference does not guarantee team making playoffs. But obviously it is fair to say that there are high correlation between them. Only Sacramento Kings, Utah Jazz, and Miami Heat were the exceptions.

Then what about 15 playoff? Usually the vibes of regular season and playoff are totally different. Many championship contender teams do not take their regular season seriously. Their real seasons start when the regular season ends.

```
po15_rating <- po15 %>%
  mutate(OffRating = PTS * (100 / (FGA - ORB + TOV + (FTA * 0.44))),
         DefRating = po15_opp$PTS * (100 / (po15_opp$FGA - po15_opp$OREB + po15_opp$TOV +(po15_opp$FTA * 0.
po15_rating <- po15_rating[c(2,26,27)] #taking team names, offratings, defratings
po15_rating <- po15_rating %>%
  mutate(diff = OffRating - DefRating)
#sort pure efficiency difference by descending order
po15_rating <- po15_rating[order(po15_rating$diff, decreasing = TRUE),]
po15_rating
```

```
##                        Team OffRating DefRating        diff
## 1   Golden State Warriors 106.44876  98.39252   8.0562459
## 10          Brooklyn Nets  99.14237  97.41900   1.7233633
## 13 Portland Trail Blazers  99.04545  98.07210   0.9733449
## 14   New Orleans Pelicans 104.09133 103.64364   0.4476895
## 9       San Antonio Spurs 106.73428 106.85557  -0.1212870
## 3         Houston Rockets 103.92414 104.42616  -0.5020237
## 6           Chicago Bulls 100.47797 101.04890  -0.5709362
## 7       Memphis Grizzlies 100.19231 101.99351  -1.8012018
## 8      Washington Wizards 103.34465 105.27655  -1.9318995
## 11       Dallas Mavericks 106.08749 108.56339  -2.4758956
## 16         Boston Celtics  97.20940 101.00815  -3.7987549
## 4           Atlanta Hawks 101.04134 106.14502  -5.1036711
## 5     Los Angeles Clippers 105.91257 111.64152  -5.7289405
## 2     Cleveland Cavaliers 103.95229 112.45410  -8.5018134
## 12        Milwaukee Bucks  90.00000 100.31255 -10.3125528
## 15        Toronto Raptors  95.43877 110.21619 -14.7774229
```

```
ggplot(data = po15_rating, aes(x = reorder(Team, diff), y = diff)) +
  geom_bar(stat = 'identity') +
  labs(x = "Team", y = "Total Efficiency") +
  ggtitle("NBA Teams ranked by Efficiency Difference in 15 Playoff") +
  geom_hline(yintercept = mean(po15_rating$diff), color = "red", size = 1) +
  coord_flip()
```



Looking back 15 NBA playoff, final four were Warriors, Cavaliers, Hawks and Rockets. But, the only team that overlaps between the plot and top four is Golden State Warriors. Then, if I want to say that the most efficient team in the season is Warrior who won the league, there is a possibility that this cannot be true because suppose one team wins 2 games in really big margins and lose 4 games. I know this is kind of unlikely but still, statistically, this does not guarantee the result.

Calculating each efficiency rating looks very similar, so I tried to make a function to make it simple.

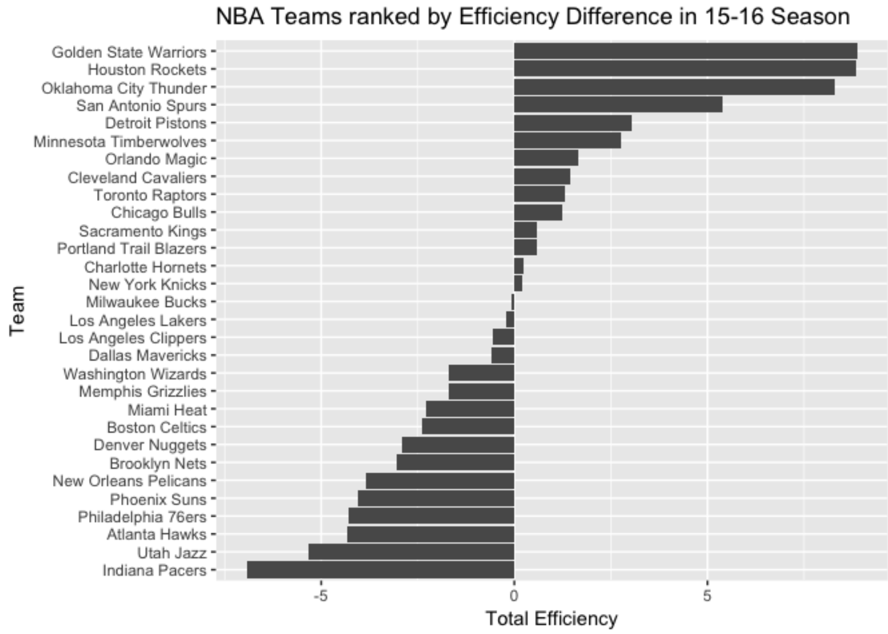
```
season_ratings <- function(season, season_opp){
  r <- season%>%
    mutate(OffRating = PTS * (100 / (FGA - ORB + TOV + (FTA * 0.44))),
         DefRating = season_opp$PTS * (100 / (season_opp$FGA - season_opp$OREB + season_opp$TOV +(season_op
  r <- r[c(2,26,27)]
  r <- r %>%
    mutate(diff = OffRating - DefRating)
  r <- r[order(r$diff, decreasing = TRUE),]
  r
}
```

So, I can make the tables really easy with using this function

```
reg15_rating <- season_ratings(reg15, reg15_opp)
reg16_rating <- season_ratings(reg16, reg16_opp)
po16_rating <- season_ratings(po16, po16_opp)
po17_rating <- season_ratings(po17, po17_opp)
```

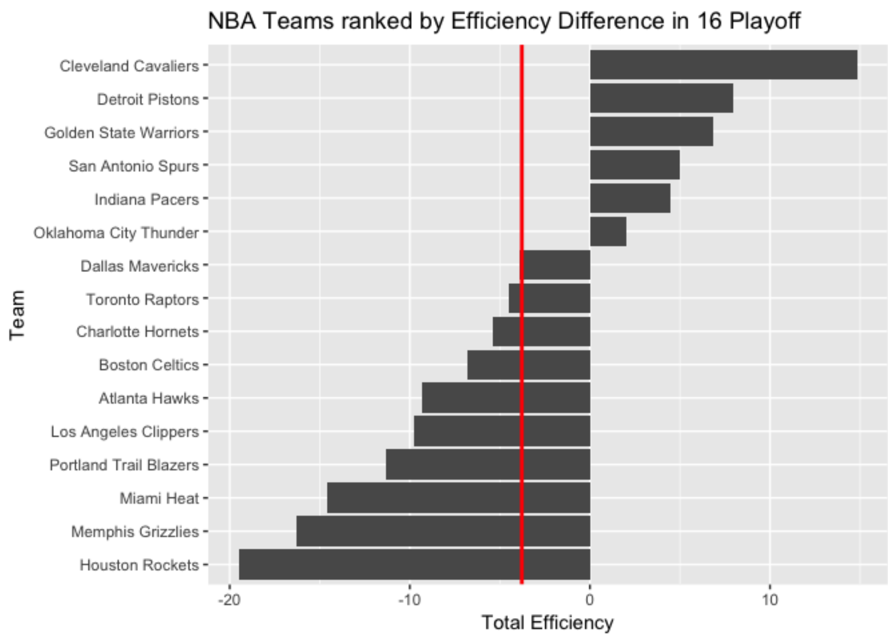
We are repeating the same process that we did above because only single season does not mean anything.

```
ggplot(data = reg15_rating, aes(x = reorder(Team, diff), y = diff)) +
  geom_bar(stat = 'identity') +
  labs(x = "Team", y = "Total Efficiency") +
  ggtitle("NBA Teams ranked by Efficiency Difference in 15-16 Season") +
  coord_flip()
```

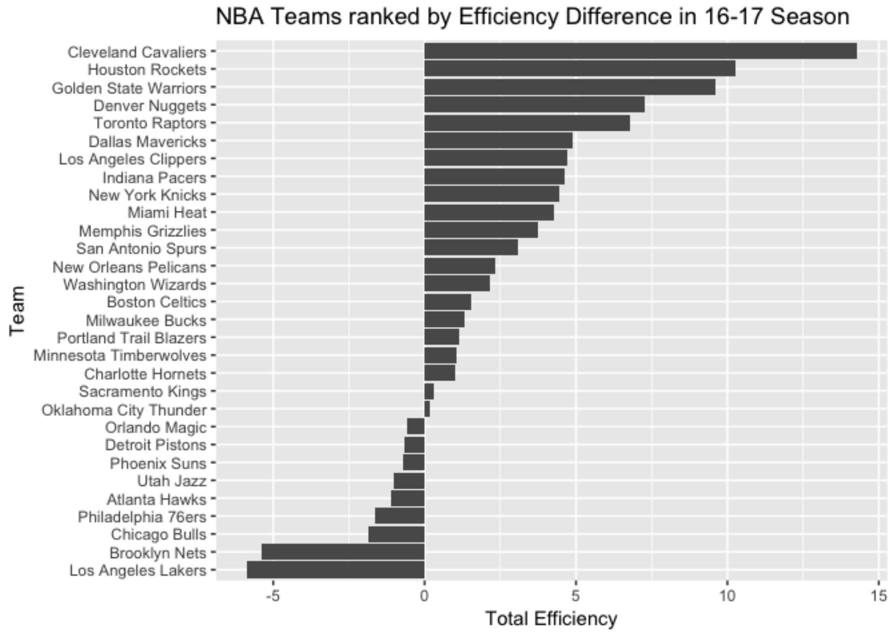


Here, I remember the last season. Houston Rockets did really well on regular season but Golden State Warriors beat them on semi-final. Oklahoma City Thunder played well too, they got beaten by Warriors in playoff though. The weird thing is, why are Detroit Pistons on top 5 even though they did not make a playoff? This means efficiency does not guarantee anything but some correaltions between efficiency and winnings.

```
ggplot(data = po16_rating, aes(x = reorder(Team, diff), y = diff)) +
  geom_bar(stat = 'identity') +
  labs(x = "Team", y = "Total Efficiency") +
  ggtitle("NBA Teams ranked by Efficiency Difference in 16 Playoff") +
  geom_hline(yintercept = mean(po16_rating$diff), color = "red", size = 1) +
  coord_flip()
```

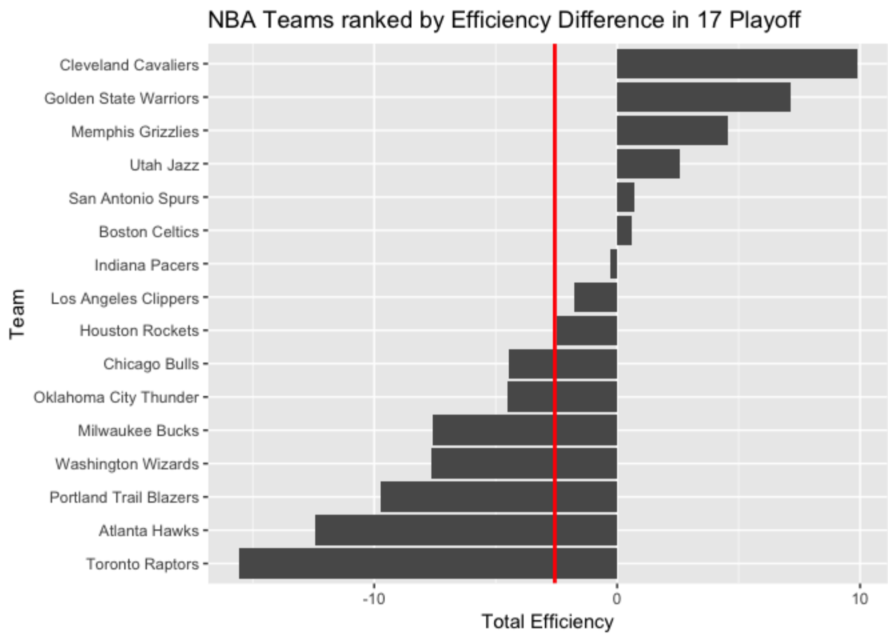



```

Cavalier won 16 playoff and it looks like the first spot does match with efficiency. Other than that, nothing is guaranteed.

```
ggplot(data = reg16_rating, aes(x = reorder(Team, diff), y = diff)) +
  geom_bar(stat = 'identity') +
  labs(x = "Team", y = "Total Efficiency") +
  ggtitle("NBA Teams ranked by Efficiency Difference in 16-17 Season") +
  coord_flip()
```
```



Glancing the plot, Cavaliers took over the league in a big margin, but we all know that Warriors beat Cavaliers in final. And here, weird thing happens again. Top 4, Denver Nuggets have a good pure efficiency even though they did not make a playoff. Dallas Mavericks too.

```
ggplot(data = po17_rating, aes(x = reorder(Team, diff), y = diff)) +
  geom_bar(stat = 'identity') +
  labs(x = "Team", y = "Total Efficiency") +
  ggtitle("NBA Teams ranked by Efficiency Difference in 17 Playoff") +
  geom_hline(yintercept = mean(po17_rating$diff), color = "red", size = 1) +
  coord_flip()
```



From this plot, the first place is Cavalier and this is not a good picture. Because we know Warriors won the league. From this, I can assure, again, even with a good efficiency, there's no guaranteed future. Also, it seems like Cavaliers won several teams in big margins but still lost to Warriors. As I mentioned before - the unlikely result just happened!

Conclusion and take home message

In a nutshell, team standings does not follow the efficiency ratings. What we have seen so far, the statistics do not guarantee any of results. In other words, this cannot be causality. Winning teams have better team efficiency stats than losing teams for sure, but we cannot say the more efficient, the better team. Likewise, statistics have this loophole. This is the reason why we should be careful setting causality between two aspects. In order to confirm the causality, sometimes we need p-value and hypothesis test. Next time, I will introduce p-value and hypothesis test.

source

1> https://stats.nba.com/teams/opponent/?sort=W&dir=-1&Season=2014-15&SeasonType=Regular%20Season&PerMode=Totals

2> https://stats.nba.com/teams/opponent/?sort=W&dir=-1&Season=2014-15&SeasonType=Playoffs&PerMode=Totals

3> https://stats.nba.com/teams/opponent/?sort=W&dir=-1&Season=2015-16&SeasonType=Regular%20Season&PerMode=Totals

4> https://stats.nba.com/teams/opponent/?sort=W&dir=-1&Season=2015-16&SeasonType=Playoffs&PerMode=Totals

5> https://stats.nba.com/teams/opponent/?sort=W&dir=-1&Season=2016-17&SeasonType=Regular%20Season&PerMode=Totals

6> https://stats.nba.com/teams/opponent/?sort=W&dir=-1&Season=2016-17&SeasonType=Playoffs&PerMode=Totals

7> https://www.sportingcharts.com/nba/stats/team-offensive-efficiency-rating/2016/

8> https://www.sportingcharts.com/nba/stats/team-defensive-efficiency-rating/2016/

9> https://www.statmethods.net/management/merging.html

10> https://cran.r-project.org/web/packages/ggplot2/vignettes/ggplot2-specs.html

11> https://www3.nd.edu/~steve/computing_with_data/11_geom_examples/ggplot_examples.html