# National Basketball Association (NBA) and 3-pointers

Final Project, DECS-922: *Data Exploration*, Winter 2019

*Subeom Lee, Tianyi Peng, and Jieul Jang*

*2019-03-17*

## Contents



```
## Warning: package 'knitr' was built under R version 3.5.3
```

# 1 Introduction

In the 2018-2019 NBA regular season, we observed births of many new 3 pointers records. For example, Klay Thompson from the Golden State Warriors, made 14 3-pointers in Chicago last October. (link) James Harden, a key player from the Houston Rockets who is known for being deadly at making three pointers, created a record of second longest 30-point games streak, 32 games, in last February. This is the first record change after 1962. (link) Also, Klay Thompson mentioned earlier succeeded 10 consecutive 3-pointers in last Jan. (link)

These observations made us interested in looking this phenomenon further, especially whether they are exceptions to the normal or they represent a shifting trend in how teams play in the NBA. Thus, we would like to answer the following questions to gain a better understanding how three pointer shooting has evolved since 1986 while taking into considerations of dimensions such as teams' standings, player age, other scoring methods in the game and a few other factors.

## 1.1 Methodology

To conduct the necessary analysis, we used detailed historical data from the nbastatR package. (link) This is a comprehensive open-source package that has additional tools to analyze the NBA data. Through some initial testing, we found that not all the tools are effective at helping us understand the questions regarding the three point shooting trend in the NBA and some of the additional functions in the NBA package are duplicated. However, we found the dataset to be helpful and sufficient to conduct the proposed research.

For additional data such as histrical standings and salary information, that are not included in nbastatR package, we downloaded CSV files from Basketball-Reference.com (link)

The data from nbastatR (`dataGameLogsPlayer` & `dataGameLogsTeam`) contain all the variables available so we trimmed the data by selecting relevant variables for analysis of each question.

## 1.2 Some of the key questions that we are trying to understand are listed below:

**1.2.1 It seems that players are getting better at making 3-pointers than 10 or 20 years ago (both on average across the league and also top 3-pointer shooters vs. top 3-pointer shooters) Is it true?**

**1.2.2 Teams with more 3-pointers tend to be the better performing teams?**

**1.2.3 Are there any relationship between players' ages and 3-pointers?**

**1.2.4 Players who are good at 3-pointers are also good at 2-pointers or free throws?**

**1.2.5 Players with high salaries are good at 3-pointers?**

**1.2.6 We want to analyze whether players can drastically improve their three point shooting skills over time or the skill is rather something people are born with.**

**1.2.7 Show the 3-pointer statistics geographically based on players' hometowns. Maybe this help illustrates the different basketball playing style across different regions, both domestic and international.**

**1.2.8 We would like to explore the importance of three point shooters in a given team by measuring the share of the team's total salary over time.**

**1.2.9 What are the expected average points of 3-pointers and 2-pointers?**

**1.2.10 If the expected average point from 3-pointers is getting higher than that of 2-pointers, how should each team's strategy chang?**

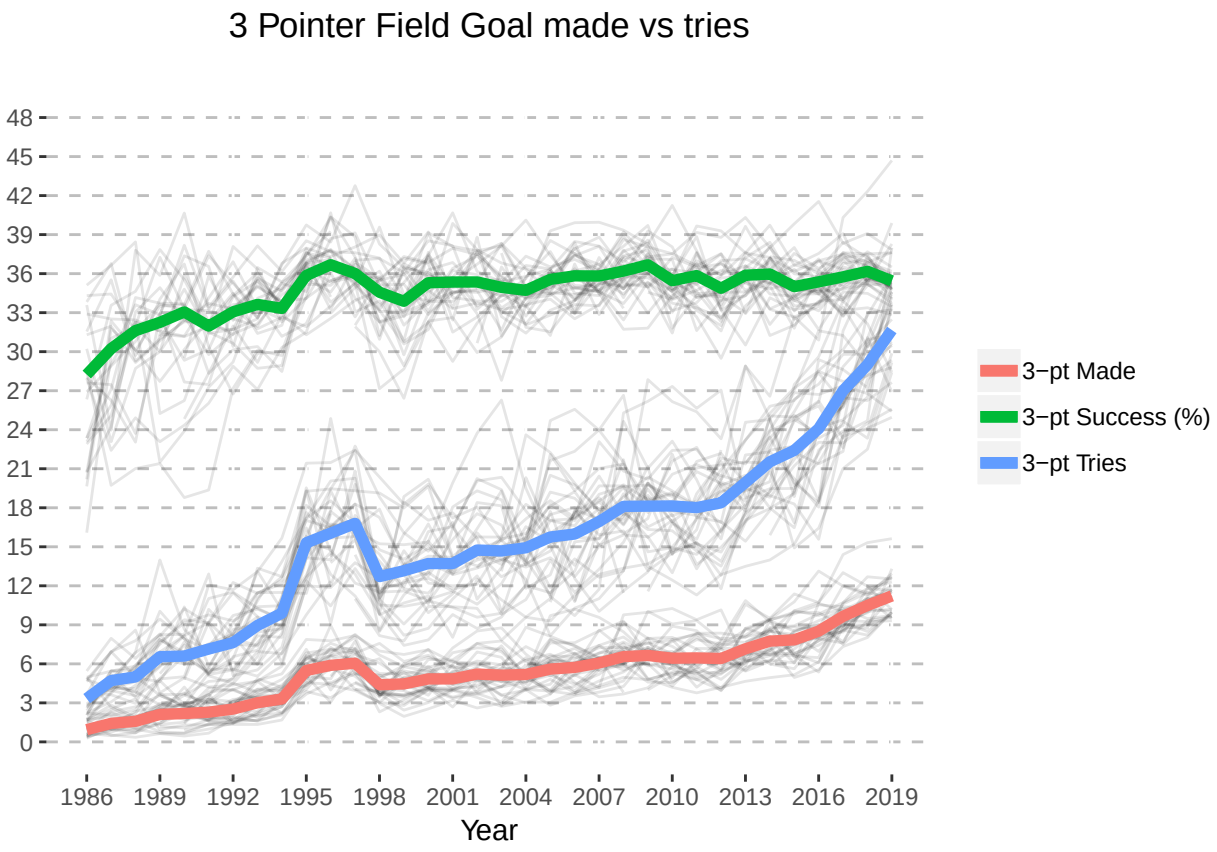
```
[1] "dataGameLogsPlayer.Rdata loaded."
[1] "dataGameLogsTeam.Rdata loaded."
[1] "df_nba_player_dict.Rdata loaded."
```

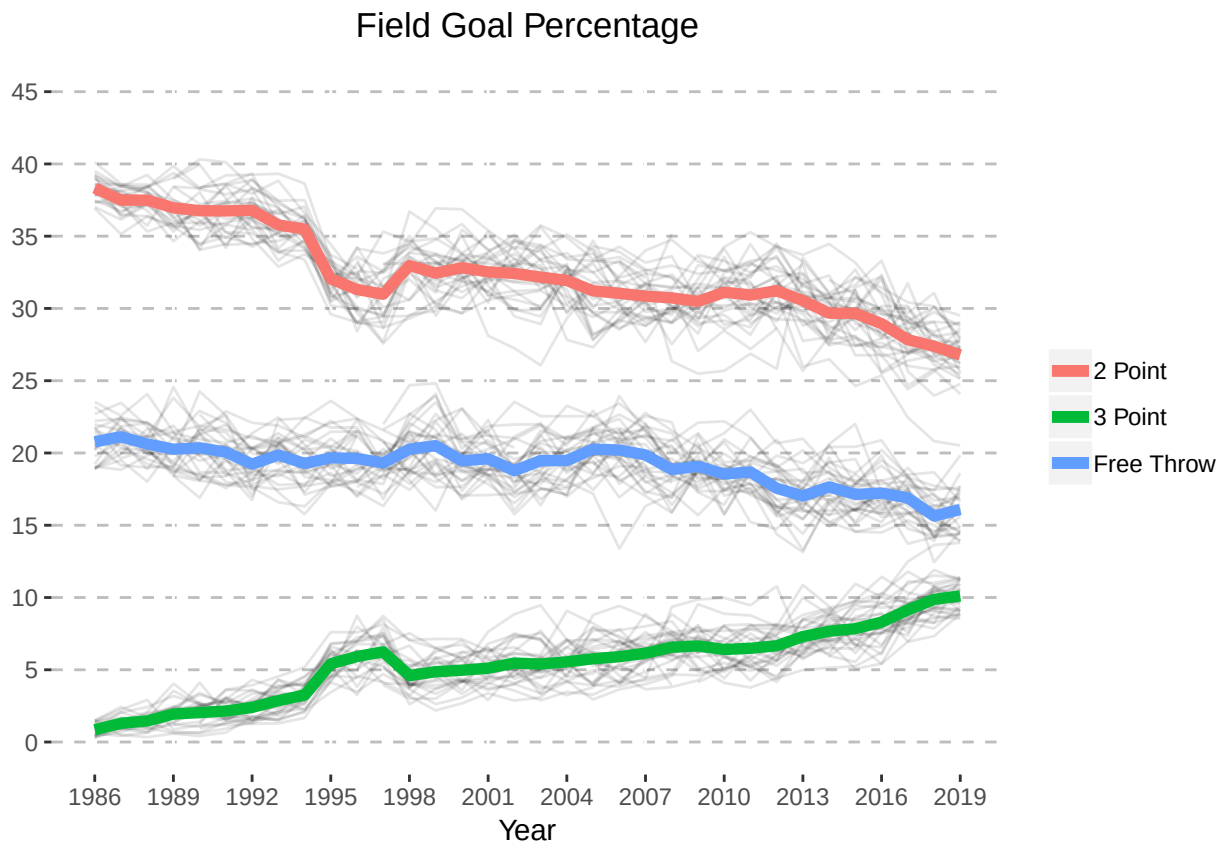
# 2 Team level questions

- **Q1. It seems that players are getting better at making 3-pointers than 20 years ago (both on average and also top 3-pointer shooters vs. top 3-pointer shooters) Is it true?**

Based on the chart 1 in the report, from 1986 to recent years, the average team in the NBA has improved their three pointer success rate, which is calculated by dividing the number of three pointers made by the number of three pointers attempted in each regular season. With that said, however, the success ratio has stayed largely flat since 1998, likely to the fact that the rules regarding three pointers have not changed and players on average have reached a plateau in terms of the three pointers success rate.

Equally interesting, instead of team average, if we were to look at the top 10 three point shooters in the league by success rate or by number of three pointers made, according to chart 2, the main driver behind their improved contribution to the total points score by their team is through the improvement of their three pointer success rate. Similar to the trend we saw among teams earlier, top three point shooters also improved their success rate over the time period that we specified here.

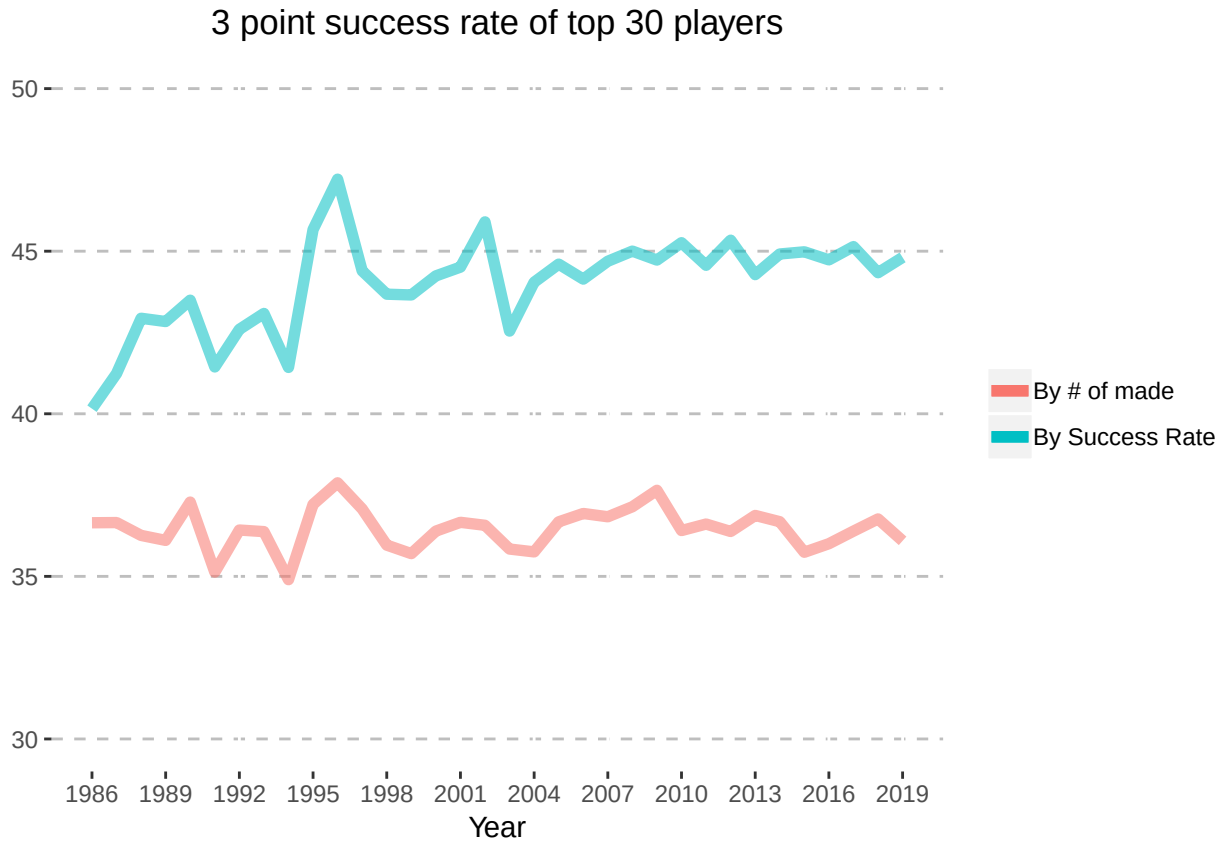





Statistics of top 10 3-point shooters each year

```
      Year         Player               fgm             fga
 Min.   :1986   Length:14714       Min.   :   0   Min.   :   0
 1st Qu.:1995   Class :character   1st Qu.:  49   1st Qu.: 115
 Median :2004   Mode  :character   Median : 154   Median : 345
 Mean   :2004                      Mean   : 200   Mean   : 437
 3rd Qu.:2012                      3rd Qu.: 308   3rd Qu.: 670
 Max.   :2019                      Max.   :1098   Max.   :2279
                                                  NA's   :9
      fg3m          fg3a            ftm           fta          pctfg3
 Min.   :  0   Min.   :  0.0   Min.   :  0   Min.   :  0   Min.   :  0.0
 1st Qu.:  0   1st Qu.:  2.0   1st Qu.: 20   1st Qu.: 29   1st Qu.: 16.0
 Median :  5   Median : 21.0   Median : 64   Median : 88   Median : 30.9
 Mean   : 29   Mean   : 82.2   Mean   :101   Mean   :134   Mean   : 26.3
 3rd Qu.: 44   3rd Qu.:128.0   3rd Qu.:145   3rd Qu.:194   3rd Qu.: 36.8
 Max.   :402   Max.   :886.0   Max.   :833   Max.   :972   Max.   :100.0
 NA's   :15    NA's   :16                    NA's   :9     NA's   :2300
     pctfg2          pctft
 Min.   :  0.0   Min.   :  0.0
 1st Qu.: 40.3   1st Qu.: 66.6
 Median : 44.3   Median : 75.0
 Mean   : 44.0   Mean   : 72.4
 3rd Qu.: 48.5   3rd Qu.: 81.4
 Max.   :100.0   Max.   :100.0
 NA's   :63      NA's   :470
# A tibble: 340 x 12
# Groups:   Year [34]
    Year Player   fgm   fga  fg3m  fg3a   ftm   fta pctfg3 pctfg2 pctft
   <int> <chr>  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  <dbl>  <dbl> <dbl>
 1  1986 Craig~   284   568    73   161    75    86   45.3   50    87.2
 2  1986 Trent~   349   739    41    91    79   100   45.1   47.2  79
 3  1986 Larry~   796  1606    82   194   441   492   42.3   49.6  89.6
```

```
 4  1986 World~   652  1428    71   169   379   486   42.0   45.7  78.0
 5  1986 Kyle ~   286   592    58   140    73    90   41.4   48.3  81.1
 6  1986 Micha~   274   606    63   163   147   170   38.7   45.2  86.5
 7  1986 Leon ~   184   463    41   112   123   155   36.6   39.7  79.4
 8  1986 Dale ~   193   470    63   174    59    82   36.2   41.1  72.0
 9  1986 Mike ~   252   544    41   114    42    64   36.0   46.3  65.6
10  1986 Brad ~   267   502    32    89   198   228   36.0   53.2  86.8
# ... with 330 more rows, and 1 more variable: Rank <int>
# A tibble: 4,651 x 12
# Groups:   Year [34]
    Year Player   fgm   fga  fg3m  fg3a   ftm   fta pctfg3 pctfg2 pctft
   <int> <chr>  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  <dbl>  <dbl> <dbl>
 1  1986 Larry~   796  1606    82   194   441   492   42.3   49.6  89.6
 2  1986 Craig~   284   568    73   161    75    86   45.3   50    87.2
 3  1986 World~   652  1428    71   169   379   486   42.0   45.7  78.0
 4  1986 Dale ~   193   470    63   174    59    82   36.2   41.1  72.0
 5  1986 Micha~   274   606    63   163   147   170   38.7   45.2  86.5
 6  1986 Kyle ~   286   592    58   140    73    90   41.4   48.3  81.1
 7  1986 John ~   365   818    45   146   231   297   30.8   44.6  77.8
 8  1986 Norm ~   403   921    42   121   131   162   34.7   43.8  80.9
 9  1986 Leon ~   184   463    41   112   123   155   36.6   39.7  79.4
10  1986 Mike ~   252   544    41   114    42    64   36.0   46.3  65.6
# ... with 4,641 more rows, and 1 more variable: Rank <int>
```




Yes, the success rate of 3-point field goal has been increased by about 9% since 1986.

- **Q2. If true, what could be the reasons for that?**
  - What are the expected average points of 3-pointers and 2-pointers? Show the historical data.
  - If the expected average point from 3-pointers is getting higher than that of 2-pointers, how should each team's strategy changes

https://www.nytimes.com/2016/01/21/sports/basketball/how-the-nba-3-point-shot-went-from-gimmick-to-game-changer.html

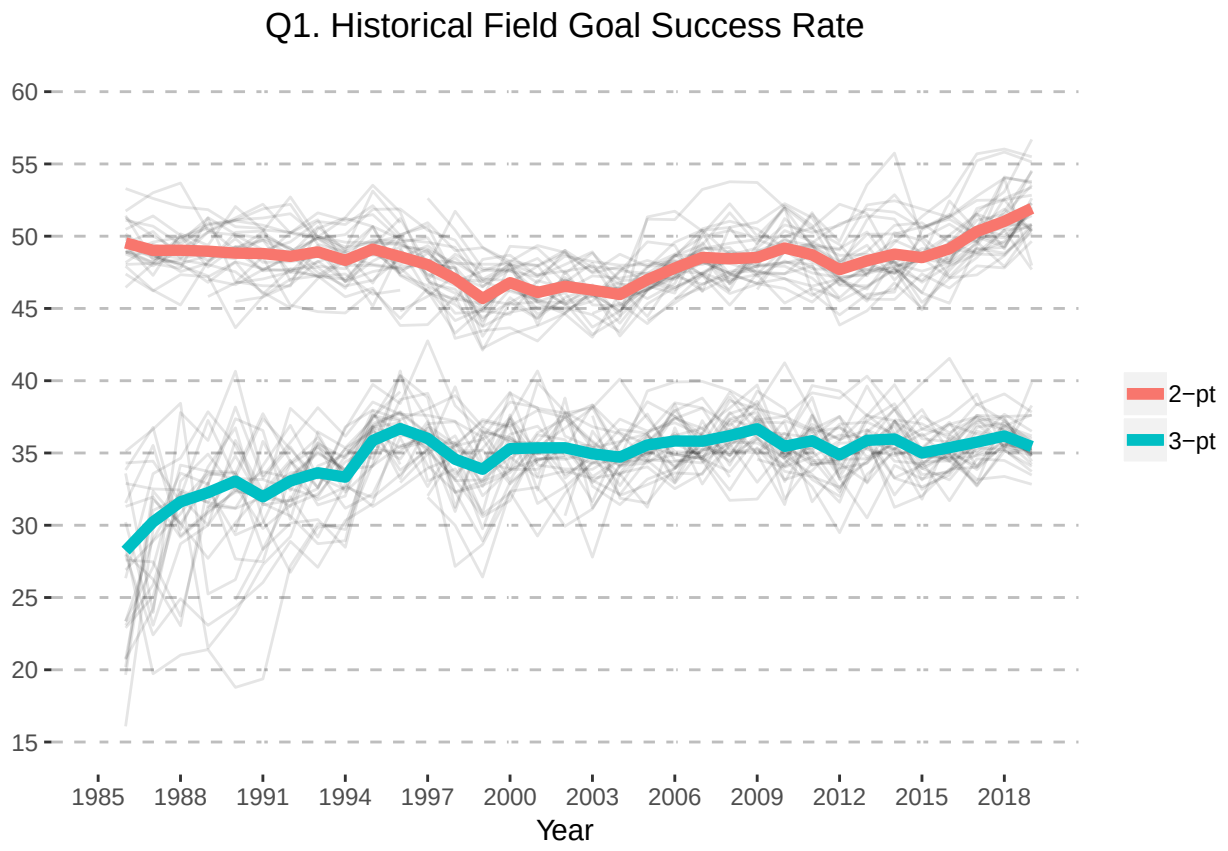
Its debut, in the 1979-80 season, was inauspicious.

There are many reasons for the rise of the 3-point shot, but one may simply be math. It took a while, but coaches finally stopped listening to the traditionalist naysayers and realized that a shot that is worth 50 percent more pays off, even if that shot is a little harder to make.

"Teams have all caught on to the whole points-per-possession argument," Lawrence Frank, the Nets' coach at the time, said in 2009 as the 3 rate began to rapidly increase.

We ran regressions on the team standings during the regular seasons by three pointer success rate and two pointer success rate. If team has a lower standings, in another word, smaller number for its ranking position, it is ranked closer to the top so the better team. Unsurprisingly, the regressions showed that independent variables three pointer success rate and two pointer success rate all have negative coefficients and they are statistically significant at 5%. This means when the three pointer and two pointer success rates increase, the teams tend to have smaller numbers as their rankings, i.e., better rankings.

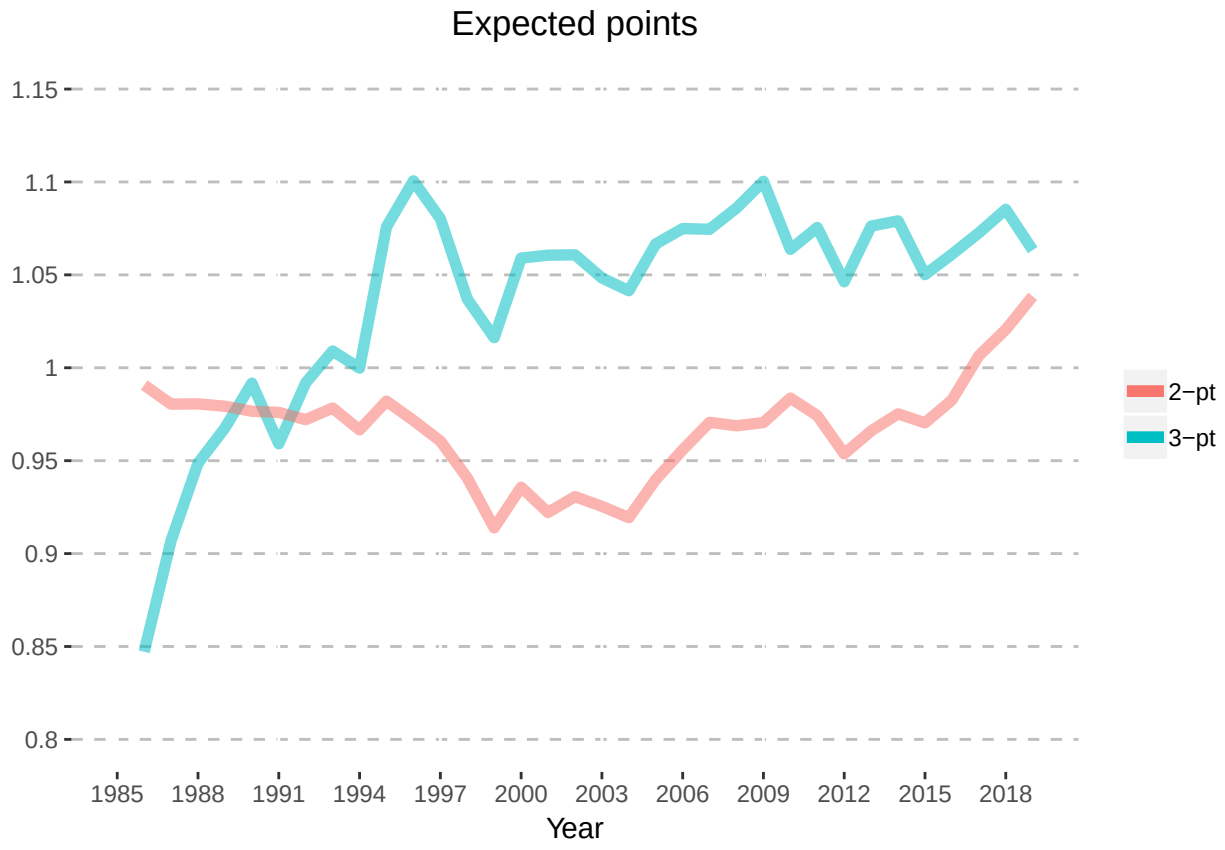
However, as two pointers still contributed to a larger share of the total points scored by a given team than three pointers do. The absolute value of the coefficient on two point success rate is larger than that of three point success rate. This is not surprising because improving the same percentage of success rate in two point shooting and three point shooting means that the former will have a larger impact on improving the team's ranking than the latter does.



The expected points of 2-point shots in 1986 was 'r fgyear$pctfg2[1986-1985]/100' * 2 =' rfgyear$pctfg2[1986-1985]/100*2' *The expected points of 3-point shots in 1986 was 'r fgyear$pctfg3[1986 - 1985]/100' * 3 =' rfgyear$pctfg3[1986-1985]/100*3'

The expected points of 2-point shots in 2019 was 'r fgyear$pctfg2[2019-1985]/100' * 2 =' rfgyear$pctfg2[2019-1985]/100*2' *The expected points of 3-point shots in 2019 was 'r fgyear$pctfg3[2019 - 1985]/100' * 3 =' rfgyear$pctfg3[2019-1985]/100*3'

Teams started to focus on 3-point shots after its first introduction in 1979, because the expected points of 3-point shots are higher than that of 2-point shots since early 90's.



- **Q3. Are there any relationship between players' ages and 3-pointers?**

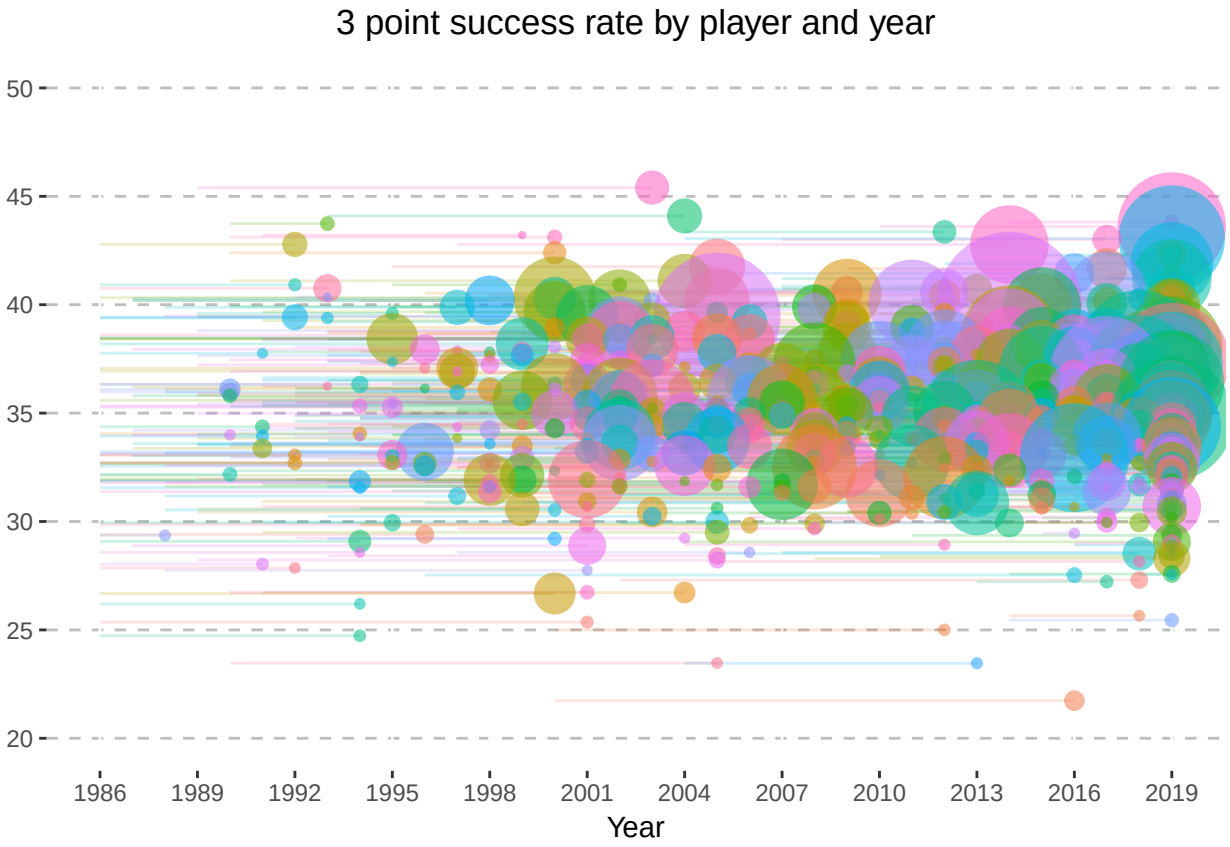
Based on the regression that we ran the three pointer accuracy on a given player's tenure in the NBA, The coefficient is positive and statistically significant at 5%, so we believe as players play longer in the NBA, they tend to have better success rate at shooting three pointers, likely due to continuing training that help improve their three pointer shooting skills and also better understanding of the games and thus know when are the right times to shoot three pointers.

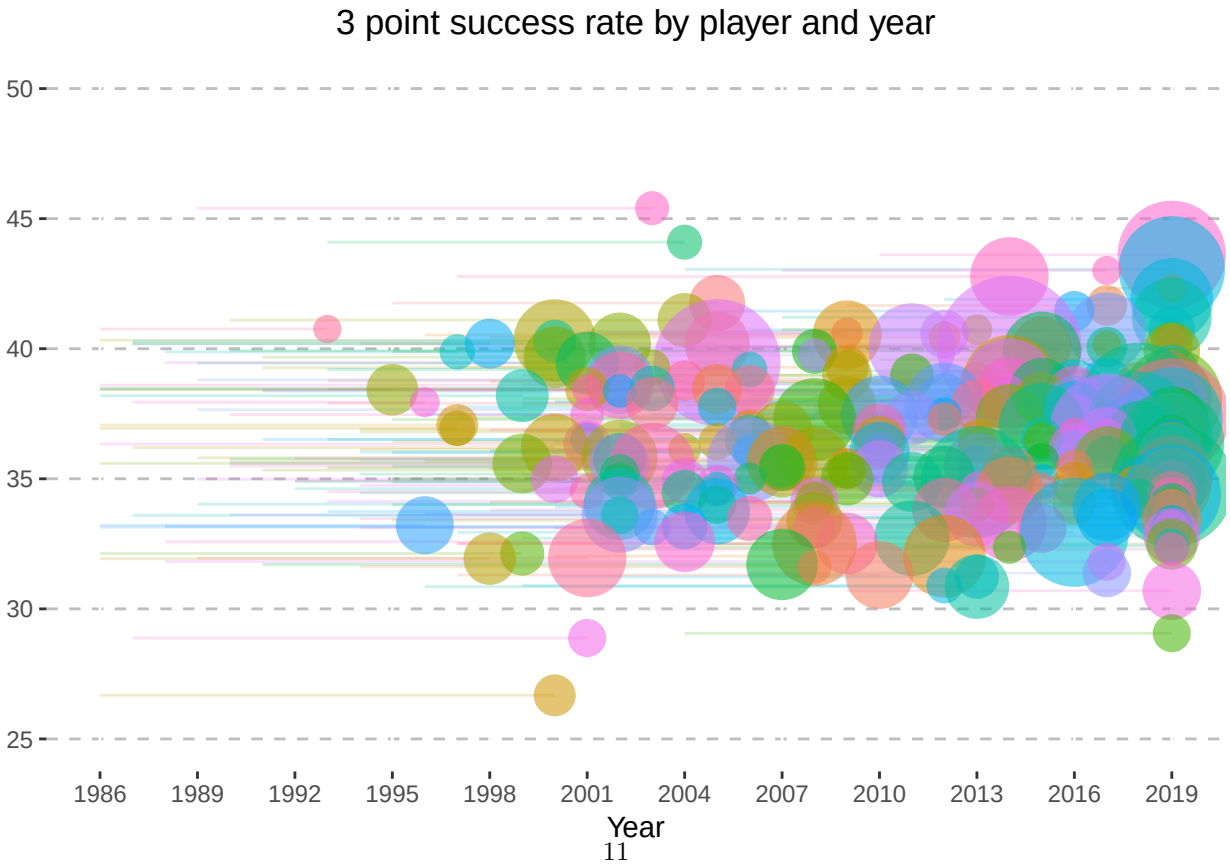
```
# A tibble: 2 x 5
  term        estimate std.error statistic  p.value
  <chr>          <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)   32.6      2.72       12.0  5.33e-31
2 pctfg3        -0.518    0.0787     -6.58 7.74e-11
# A tibble: 2 x 5
  term        estimate std.error statistic  p.value
  <chr>          <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)   107.       4.97       21.6 2.14e-84
2 pctfg2         -1.91     0.103     -18.6 3.69e-66
# A tibble: 3 x 5
  term        estimate std.error statistic  p.value
  <chr>          <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)  114.       5.15       22.1  9.52e-88
2 pctfg3        -0.305    0.0694     -4.40 1.23e- 5
3 pctfg2        -1.83     0.103     -17.7  4.80e-61
# A tibble: 2 x 5
  term        estimate std.error statistic  p.value
  <chr>          <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)   22.0      2.29        9.60 6.40e-21
2 pctfg2         0.257    0.0472      5.45 6.57e- 8
```

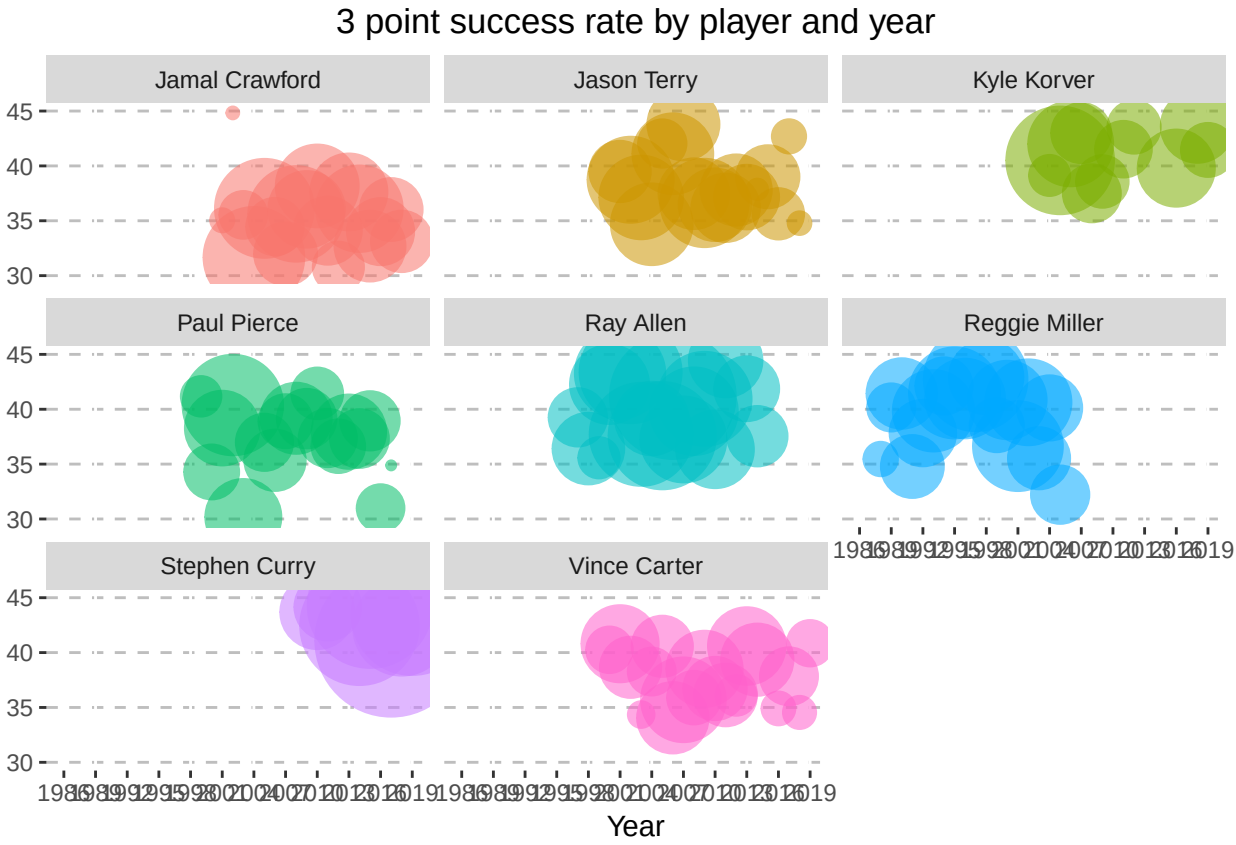
Yes. However, pctfg2 is more relevant than pctfg3

- Focus on three point shooting is a strategy that started fairly recently, we can create a map to show where this strategy initially emerged and how fast it spreaded across the entire country.

## 3 Player level questions






Above graph shows more players are trying 3 point shots than before. even though the average success rate is similar.

- **Q4. Players who are good at 3-pointers are also good at 2-pointers or free throws?**

We ran a series of regressions and found that three pointer success rate has a positive and statistically significant relationship with free throw success rate, however, we are not able to draw such conclusions between the three pointer success rate and two pointer success rate.

A couple of factors that could have contributed to this observation. Players in NBA tend to specialize in three pointer shooting vs two pointer shooting. These two specialization often mean that they need to make a trade off between which style of play to focus on at each other's expense. Also, two pointer plays include lay ups and dunks that usually top three point shooters would not do because they tend to play in the peripheral of the court during the game. One can also argue in the NBA, there is a trade off between explosiveness and three point shooting consistency. The explosive quality is important to have do well in the two pointer style play and thus created this lack of positive correlation between three point success rate and two point success rate.

```
# A tibble: 2 x 5
  term        estimate std.error statistic  p.value
  <chr>          <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)  33.7       1.75      19.2   2.81e-67
2 pctfg2        0.0330    0.0400     0.823 4.11e- 1
# A tibble: 2 x 5
  term        estimate std.error statistic  p.value
  <chr>          <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)  184.     19.6          9.41 6.19e-20
2 fgm            0.143   0.00618     23.1  2.24e-89
# A tibble: 2 x 5
  term        estimate std.error statistic   p.value
  <chr>          <dbl>     <dbl>     <dbl>     <dbl>
1 (Intercept)  404.     48.0          8.42 1.98e- 16
2 fga            0.197   0.00687     28.6  3.67e-122
# A tibble: 3 x 5
  term        estimate std.error statistic  p.value
  <chr>          <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)  276.      47.4         5.82 8.67e- 9
2 fga            0.347    0.0172     20.2  7.38e-73
3 fta           -0.455    0.0481     -9.47 3.52e-20
# A tibble: 2 x 5
  term        estimate std.error statistic  p.value
  <chr>          <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)   18.2      1.42        12.8 3.40e-34
2 pctft          0.216    0.0181      11.9 4.54e-30
# A tibble: 2 x 5
  term        estimate std.error statistic   p.value
  <chr>          <dbl>     <dbl>     <dbl>     <dbl>
1 (Intercept)  41.9       1.42       29.6  4.07e-128
2 pctft         0.0219    0.0180      1.21 2.25e-  1
# A tibble: 3 x 5
  term        estimate std.error statistic  p.value
  <chr>          <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)  17.7       2.10       8.42  1.86e-16
2 pctfg2        0.0136    0.0368     0.370 7.12e- 1
3 pctft         0.216     0.0182    11.9   6.51e-30
```

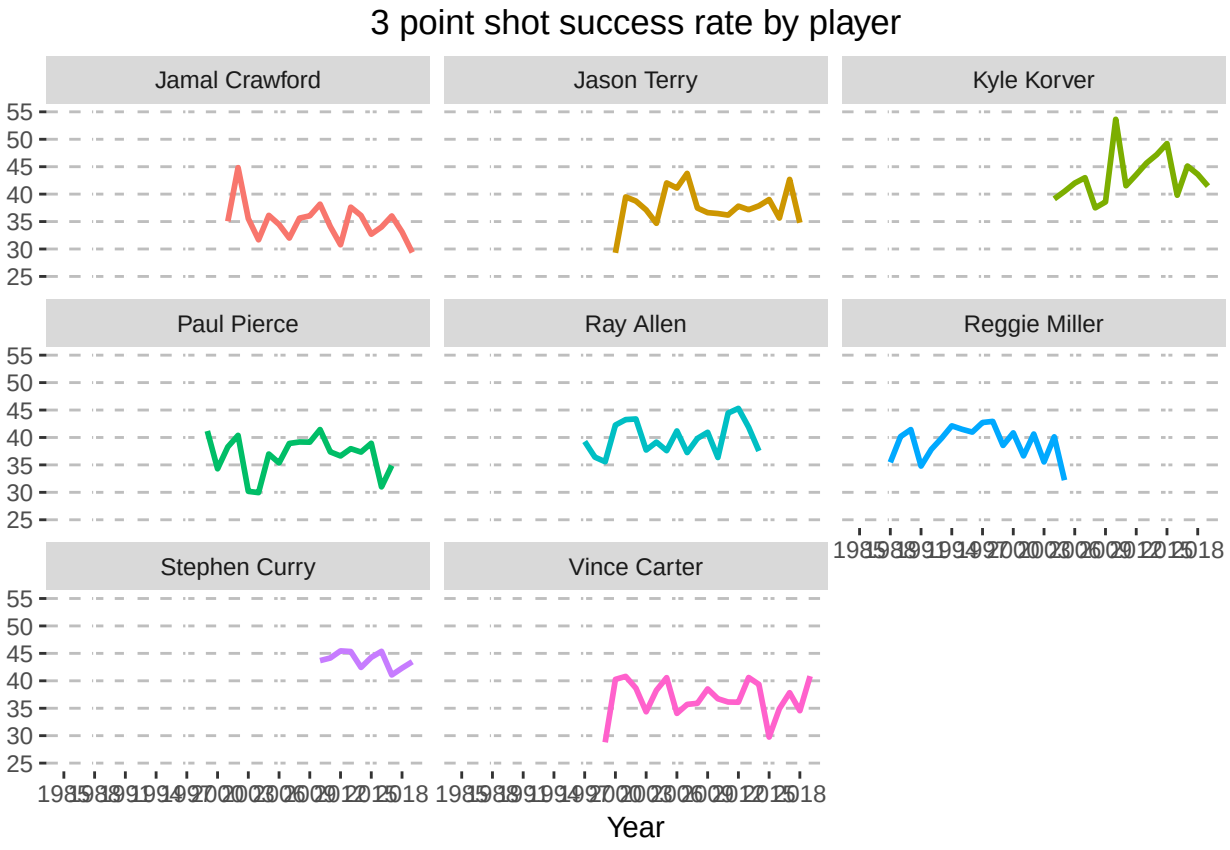
When we look at all the players, 2-pointers and 3-pointers are reverse-related. Maybe because of dunk shots?

```
# A tibble: 3 x 5
  term        estimate std.error statistic  p.value
  <chr>          <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)   3.65      2.52        1.45 1.48e- 1
2 pctfg2       -0.0441    0.0415     -1.06 2.88e- 1
3 pctft         0.329     0.0237     13.9  3.19e-42
```

Best players (more than 1,000 career 3-point field goals) are good at 2-pointers as well!!!

```
# A tibble: 3 x 5
  term        estimate std.error statistic       p.value
  <chr>          <dbl>     <dbl>     <dbl>         <dbl>
1 (Intercept)    3.76     4.06       0.926 0.356
2 pctfg2         0.345    0.0843     4.09  0.0000841
3 pctft          0.226    0.0344     6.58  0.00000000197
# A tibble: 3 x 5
  term        estimate std.error statistic p.value
  <chr>          <dbl>     <dbl>     <dbl>   <dbl>
1 (Intercept)  -21.5      20.1       -1.07   0.334
2 pctfg2         0.799     0.442      1.81   0.131
3 pctft          0.290     0.231      1.26   0.264
```

-. Are there any relationship between players' ages and 3-pointers? Both total and average.




Let's regress.

```
# A tibble: 2 x 5
  term        estimate std.error statistic  p.value
  <chr>          <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)  39.6       0.720      55.0  1.01e-95
2 career       -0.0994    0.0656     -1.51 1.32e- 1
# A tibble: 2 x 5
  term        estimate std.error statistic p.value
```

```
  <chr>          <dbl>     <dbl>     <dbl>   <dbl>
1 (Intercept)  35.4       0.281     126.    0
2 career        0.0730    0.0306      2.38  0.0173
# A tibble: 2 x 5
  term        estimate std.error statistic  p.value
  <chr>          <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)   31.7      0.208     153.   0.
2 career         0.186    0.0280      6.63 3.63e-11
# A tibble: 2 x 5
  term        estimate std.error statistic  p.value
  <chr>          <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)   24.1      0.252       95.5 0.
2 career         0.414    0.0378      11.0 7.90e-28
```

Really good players are not related with ages/career. Average players' success rate is increased by 0.4% in one year. Not bad. . . ?

- Players with high salaries are good at 3-pointers?

2018-2019 season data only

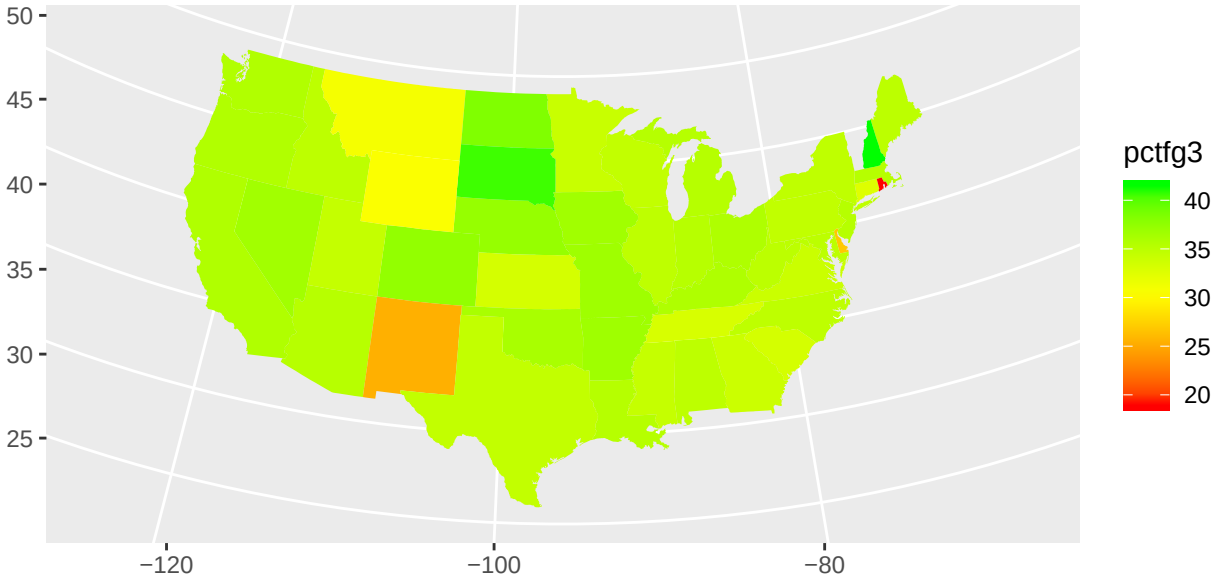
```
[1] "nbaInsiderSalaries.Rdata loaded."
# A tibble: 2 x 5
  term        estimate std.error statistic p.value
  <chr>          <dbl>     <dbl>     <dbl>   <dbl>
1 (Intercept)    1.10     25.9      0.0426   0.966
2 pctfg3         0.580     0.690    0.841    0.403
# A tibble: 2 x 5
  term        estimate std.error statistic p.value
  <chr>          <dbl>     <dbl>     <dbl>   <dbl>
1 (Intercept) 16.1       5.69         2.83 0.00587
2 fg3m         0.00460   0.00374      1.23 0.223
```

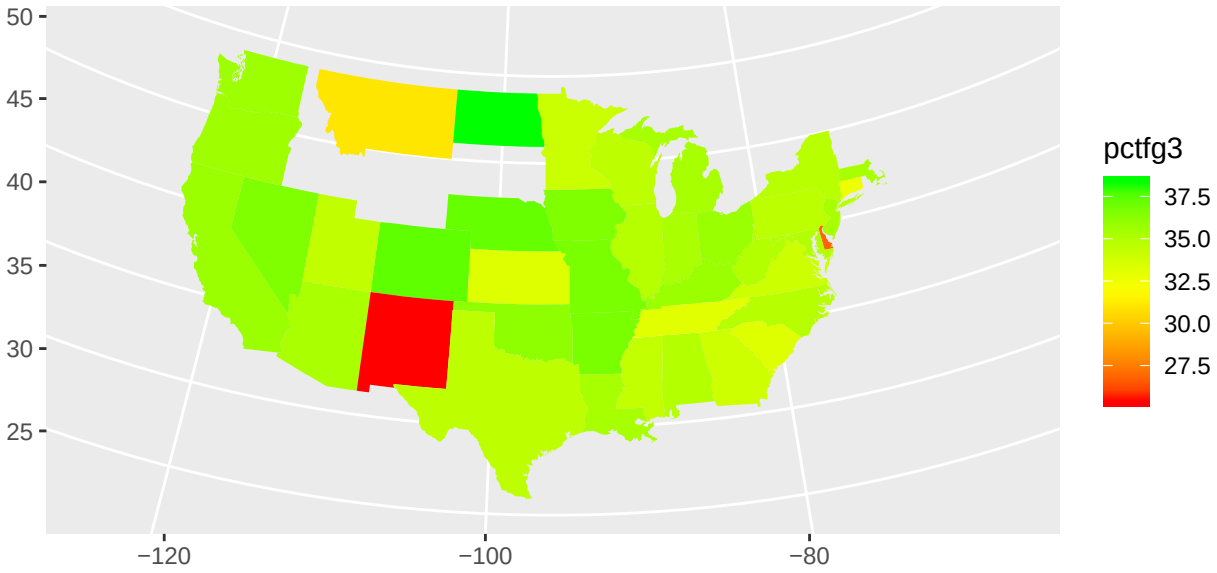
When the salary increases by a million dollar, career success rate of 3-point shots increases by 0.09% only. It's difficult to say that 3-pointer success rate is the most important factor for one's salary.

- We would like to explore the importance of three point shooters in a given team by measuring the share of the team's total salary over time.
- We want to analyze whether players can drastically improve their three point shooting skills over time or the skill is rather something people are borned with.

There is no dramatic increase in 3-pointer success rate. Maybe if we can check the players' data from NCAA or high school league, there might be different insight. However, based on NBA data, no big changes.

- Show the 3-pointer statistics geographically based on players' hometowns. Maybe this help illustrates the different basketball playing style across different regions, both domestic and international.


50
45
40
35
30
25
−120
−100
−80
pctfg3
40
35
30
25
20


50
45
40
35
30
25
−120
−100
−80
pctfg3
37.5
35.0
32.5
30.0
27.5