

Project 2 - NBA Stats

Team: Maria Ginorio, Euclides Rodriguez, Peter Gatica

03/12/2021

Introduction - NBA Leaders in Three pointers, Field Goals, and Free Throws as of March, 9, 2021 through 37 games.

I really enjoy sports statistics especially baseball and basketball stats. I chose basketball for this project because in baseball we count everything and there are just too many for my inexperience with data analytics. Basketball has some highly visible and very exciting stats like the 3 pointer and some not so exciting but just as important when a game is on the line but each can determine the outcome of a game.

Source: NBA Advanced Stats

I took a manual copy of the NBA Players stats page for the 492 players in the league as of 03/09/2021 and created a text file. When saved the file contents was tab delimited with some extraneous fields and data. Below you will find a sample.

```
filename <- "https://raw.githubusercontent.com/audiorunner13/Masters-Coursework/main/DATA607%20Spring2022/2022%20Spring%20Data%20607%20Final%20Project/nba_player_stats.csv"
nba_player_stats <- readLines(filename)
nba_player_stats[1:19]
```

```
## [1] "Season Type"
## [2] "Per Mode"
## [3] "Stat Category"
## [4] "Advanced Filters"
## [5] "Totals"
## [6] "Recent Filters Glossary Share"
## [7] "492 Rows | Page"
## [8] "of 10"
## [9] "# \t"
## [10] "Player \tGP \tMIN \tPTS \tFGM \tFGA \tFG% \t3PM \t3PA \t3P% \tFTM \tFTA \tFT% \tOREB \tDREB \tP"
## [11] "1 \t"
## [12] "Bradley Beal"
## [13] "\t33 \t1178 \t1074 \t374 \t786 \t47.6 \t75 \t229 \t32.8 \t251 \t277 \t90.6 \t43 \t134 \t177 \t"
## [14] "2 \t"
## [15] "Stephen Curry"
## [16] "\t35 \t1194 \t1039 \t343 \t718 \t47.8 \t169 \t411 \t41.1 \t184 \t197 \t93.4 \t17 \t174 \t191 \t"
## [17] "3 \t"
## [18] "Giannis Antetokounmpo"
## [19] "\t35 \t1189 \t1015 \t368 \t661 \t55.7 \t39 \t137 \t28.5 \t240 \t363 \t66.1 \t64 \t344 \t408 \t"
```

Initialize `nba_player_rows` vector to hold the rows of player data.

```
nba_player_rows <- c()
```

The first 11 rows in the dataset are extraneous and will not be used for our purposes here so I will start at row 11 and process all 1486 rows advancing each read by 3 rows.

I use the following functions to process and tidy up the nba_player_stats:

1. `unlist()` produces a vector which contains all the atomic components which occur in list (row).
2. `strsplit()` will split the elements of a character vector into substrings according to the matching separator. In this case the tab delimiter.
3. `gsub()` will search for white space and replace with no space
4. `str_to_title()` will convert every first letter in each word to a capital letter.
5. `rbind()` will combine the rows created into the vector initialized above.
6. `seq()` is used to cycle through all rows in the dataset starting at row 11, ending with the last row and advancing every 3 rows.

```
for (i in seq(12, 1486, 3)) {
  # row1 <- unlist(strsplit(nba_player_stats[i], '\\\\t'))
  # player_row <- gsub(' ', '', row1[1])

  row1 <- unlist(strsplit(nba_player_stats[i], '\\\\t'))
  player_name <- str_to_title(str_trim(row1[1]))

  row2 <- unlist(strsplit(nba_player_stats[i+1], '\\\\t'))
  games_played <- gsub(' ', '', row2[2])
  min_played <- gsub(' ', '', row2[3])
  pts_scored <- gsub(' ', '', row2[4])
  fg_made <- gsub(' ', '', row2[5])
  fg_attempted <- gsub(' ', '', row2[6])
  fg_perc <- gsub(' ', '', row2[7])
  three_pts_made <- gsub(' ', '', row2[8])
  three_pts_attempted <- gsub(' ', '', row2[9])
  three_perc <- gsub(' ', '', row2[10])
  ft_made <- gsub(' ', '', row2[11])
  ft_attempted <- gsub(' ', '', row2[12])
  ft_perc <- gsub(' ', '', row2[13])

  nba_player_rows <- rbind(nba_player_rows, c(player_name, games_played, min_played, pts_scored, fg_made,
  })
```

Next I created a data.frame from the list of concatenated rows of NBA players and only subsetting the field that I believe are the most important statistics in basketball to include 3 Point Percentage, Field Goal Percentage, and Free Throw Percentage.

```
nba_player_stats_df <- data.frame(nba_player_rows)
colnames(nba_player_stats_df) <- c("Player", "games_played", "min_played", "pts_scored", "fg_made", "fg_attempted", "fg_perc", "three_pts_made", "three_pts_attempted", "three_perc", "ft_made", "ft_attempted", "ft_perc")
head(nba_player_stats_df, 50)
```

##	Player	games_played	min_played	pts_scored	fg_made
## 1	Bradley Beal	33	1178	1074	374
## 2	Stephen Curry	35	1194	1039	343
## 3	Giannis Antetokounmpo	35	1189	1015	368

## 4	Damian Lillard	34	1224	1013	315	
## 5	Nikola Jokic	36	1293	977	379	
## 6	Zach Lavine	34	1209	976	350	
## 7	Luka Doncic	33	1164	938	328	
## 8	Lebron James	36	1245	930	350	
## 9	Joel Embiid	30	990	905	285	
## 10	Trae Young	34	1182	897	262	
## 11	Nikola Vucevic	36	1220	886	355	
## 12	Zion Williamson	34	1126	870	335	
## 13	Brandon Ingram	36	1255	861	306	
## 14	Julius Randle	37	1358	860	307	
## 15	Donovan Mitchell	34	1154	839	288	
## 16	Jaylen Brown	33	1115	815	305	
## 17	De'aaron Fox	35	1198	806	293	
## 18	Kawhi Leonard	30	1039	799	287	
## 19	James Harden	31	1173	784	250	
## 20	Jayson Tatum	31	1110	773	279	
## 21	Jerami Grant	33	1194	773	254	
## 22	Devin Booker	31	1050	772	284	
## 23	Collin Sexton	31	1127	747	274	
## 24	Khris Middleton	36	1203	741	272	
## 25	Jamal Murray	34	1204	740	273	
## 26	Domantas Sabonis	35	1278	727	274	
## 27	Malcolm Brogdon	34	1211	721	276	
## 28	Kyrie Irving	26	924	708	269	
## 29	Shai Gilgeous-Alexander	30	1020	697	239	
## 30	Fred Vanvleet	34	1249	684	226	
## 31	Malik Beasley	33	1093	677	250	
## 32	Terry Rozier	33	1106	676	245	
## 33	Gordon Hayward	31	1080	668	245	
## 34	John Collins	36	1104	647	252	
## 35	Jordan Clarkson	36	937	646	237	
## 36	Paul George	27	904	640	225	
## 37	Bam Adebayo	33	1119	634	236	
## 38	Andrew Wiggins	37	1193	629	241	
## 39	Tobias Harris	31	1066	627	241	
## 40	Norman Powell	34	1006	626	211	
## 41	Rj Barrett	37	1227	612	233	
## 42	Demar Derozan	30	1018	609	206	
## 43	Pascal Siakam	30	1075	602	222	
## 44	Buddy Hield	36	1253	590	198	
## 45	Russell Westbrook	28	958	567	221	
## 46	Bojan Bogdanovic	36	1102	566	188	
## 47	Lamelo Ball	35	1010	553	201	
## 48	Coby White	34	1130	552	201	
## 49	Harrison Barnes	33	1179	552	184	
## 50	Kevin Durant	19	680	551	188	
##	fg_attempts	fg_pct	three_pts_made	three_pts_attempts	three_pts_pct	ft_made
## 1	786	47.6	75	229	32.8	251
## 2	718	47.8	169	411	41.1	184
## 3	661	55.7	39	137	28.5	240
## 4	708	44.5	146	380	38.4	237
## 5	668	56.7	56	134	41.8	163
## 6	667	52.5	120	276	43.5	156

## 7	688	47.7	87	248	35.1	195
## 8	687	50.9	86	240	35.8	144
## 9	547	52.1	37	89	41.6	298
## 10	610	43.0	85	225	37.8	288
## 11	735	48.3	96	233	41.2	80
## 12	546	61.4	6	19	31.6	194
## 13	648	47.2	85	224	37.9	164
## 14	636	48.3	69	169	40.8	177
## 15	684	42.1	113	296	38.2	150
## 16	616	49.5	76	196	38.8	129
## 17	629	46.6	64	187	34.2	156
## 18	562	51.1	60	155	38.7	165
## 19	517	48.4	101	252	40.1	183
## 20	638	43.7	81	220	36.8	134
## 21	586	43.3	81	223	36.3	184
## 22	573	49.6	63	174	36.2	141
## 23	563	48.7	52	130	40.0	147
## 24	549	49.5	83	192	43.2	114
## 25	561	48.7	90	226	39.8	104
## 26	522	52.5	32	96	33.3	147
## 27	617	44.7	82	217	37.8	87
## 28	527	51.0	76	185	41.1	94
## 29	468	51.1	61	148	41.2	158
## 30	565	40.0	114	298	38.3	118
## 31	550	45.5	116	286	40.6	61
## 32	500	49.0	111	253	43.9	75
## 33	503	48.7	64	158	40.5	114
## 34	470	53.6	50	129	38.8	93
## 35	530	44.7	114	308	37.0	58
## 36	451	49.9	95	211	45.0	95
## 37	419	56.3	2	6	33.3	160
## 38	521	46.3	65	185	35.1	82
## 39	470	51.3	51	127	40.2	94
## 40	432	48.8	88	201	43.8	116
## 41	526	44.3	43	123	35.0	103
## 42	420	49.0	17	53	32.1	180
## 43	490	45.3	36	122	29.5	122
## 44	507	39.1	140	370	37.8	54
## 45	512	43.2	30	105	28.6	95
## 46	435	43.2	98	242	40.5	92
## 47	448	44.9	70	185	37.8	81
## 48	479	42.0	81	229	35.4	69
## 49	374	49.2	56	143	39.2	128
## 50	359	52.4	49	113	43.4	126
##	ft_attemps	ft_pct				
## 1	277	90.6				
## 2	197	93.4				
## 3	363	66.1				
## 4	254	93.3				
## 5	185	88.1				
## 6	182	85.7				
## 7	258	75.6				
## 8	208	69.2				
## 9	348	85.6				

## 10	328	87.8
## 11	94	85.1
## 12	279	69.5
## 13	185	88.6
## 14	220	80.5
## 15	182	82.4
## 16	167	77.2
## 17	232	67.2
## 18	189	87.3
## 19	213	85.9
## 20	156	85.9
## 21	207	88.9
## 22	168	83.9
## 23	183	80.3
## 24	129	88.4
## 25	123	84.6
## 26	206	71.4
## 27	97	89.7
## 28	106	88.7
## 29	199	79.4
## 30	132	89.4
## 31	72	84.7
## 32	91	82.4
## 33	133	85.7
## 34	109	85.3
## 35	60	96.7
## 36	107	88.8
## 37	189	84.7
## 38	122	67.2
## 39	106	88.7
## 40	129	89.9
## 41	141	73.0
## 42	204	88.2
## 43	149	81.9
## 44	64	84.4
## 45	164	57.9
## 46	108	85.2
## 47	101	80.2
## 48	77	89.6
## 49	155	82.6
## 50	145	86.9

I am shortening the name of Giannis Antetokounmpo to G. Antetokounmpo for display purposes as I noticed that his name would overrun into the next players label in any graph his name was present.

```
nba_player_stats_df$Player[nba_player_stats_df$Player == "Giannis Antetokounmpo"] <- "G. Antetokounmpo"
```

Transformation of character fields to integers and doubles.

```
nba_player_stats_df$games_played <- as.integer(nba_player_stats_df$games_played)
nba_player_stats_df$min_played <- as.integer(nba_player_stats_df$min_played)
nba_player_stats_df$pts_scored <- as.integer(nba_player_stats_df$pts_scored)
nba_player_stats_df$fg_made <- as.integer(nba_player_stats_df$fg_made)
```

```

nba_player_stats_df$fg_attempts <- as.integer(nba_player_stats_df$fg_attempts)
nba_player_stats_df$fg_pct <- as.double(nba_player_stats_df$fg_pct)
nba_player_stats_df$three_pts_made <- as.integer(nba_player_stats_df$three_pts_made)
nba_player_stats_df$three_pts_attempts <- as.integer(nba_player_stats_df$three_pts_attempts)
nba_player_stats_df$three_pts_pct <- as.double(nba_player_stats_df$three_pts_pct)
nba_player_stats_df$ft_made <- as.integer(nba_player_stats_df$ft_made)
nba_player_stats_df$ft_attempts <- as.integer(nba_player_stats_df$ft_attempts)
nba_player_stats_df$ft_pct <- as.double(nba_player_stats_df$ft_pct)

head(nba_player_stats_df,10)

```

```

##           Player games_played min_played pts_scored fg_made fg_attempts
## 1      Bradley Beal          33       1178       1074      374        786
## 2      Stephen Curry          35       1194       1039      343        718
## 3  G. Antetokounmpo          35       1189       1015      368        661
## 4      Damian Lillard          34       1224       1013      315        708
## 5      Nikola Jokic           36       1293        977      379        668
## 6      Zach Lavine            34       1209        976      350        667
## 7      Luka Doncic            33       1164        938      328        688
## 8      LeBron James           36       1245        930      350        687
## 9      Joel Embiid            30        990        905      285        547
## 10     Trae Young             34       1182        897      262        610
##      fg_pct three_pts_made three_pts_attempts three_pts_pct ft_made ft_attempts
## 1      47.6             75             229          32.8      251        277
## 2      47.8            169             411          41.1      184        197
## 3      55.7             39             137          28.5      240        363
## 4      44.5            146             380          38.4      237        254
## 5      56.7             56             134          41.8      163        185
## 6      52.5            120             276          43.5      156        182
## 7      47.7             87             248          35.1      195        258
## 8      50.9             86             240          35.8      144        208
## 9      52.1             37              89          41.6      298        348
## 10     43.0             85             225          37.8      288        328
##      ft_pct
## 1      90.6
## 2      93.4
## 3      66.1
## 4      93.3
## 5      88.1
## 6      85.7
## 7      75.6
## 8      69.2
## 9      85.6
## 10     87.8

```

```
summary(nba_player_stats_df)
```

```

##      Player      games_played      min_played      pts_scored
## Length:492      Min.   : 1.00      Min.   : 4.0      Min.   : 0.0
## Class :character 1st Qu.:15.00      1st Qu.: 153.2      1st Qu.: 51.0
## Mode  :character Median :25.00      Median : 524.5      Median : 186.5
##                      Mean  :23.07      Mean   : 525.3      Mean   : 243.8

```

```
##           3rd Qu.:33.00   3rd Qu.: 836.2   3rd Qu.: 364.5
##           Max.      :38.00   Max.      :1358.0   Max.      :1074.0
##   fg_made   fg_attempts   fg_pct   three_pts_made
## Min.      : 0.00   Min.      : 0.00   Min.      : 0.00   Min.      : 0.00
## 1st Qu.: 19.00   1st Qu.: 44.75   1st Qu.: 39.10   1st Qu.: 3.00
## Median : 69.00   Median :148.00   Median : 44.20   Median : 17.00
## Mean      : 89.35   Mean      :192.17   Mean      : 43.68   Mean      : 27.91
## 3rd Qu.:135.25   3rd Qu.:283.75   3rd Qu.: 50.00   3rd Qu.: 46.00
## Max.      :379.00   Max.      :786.00   Max.      :100.00   Max.      :169.00
## three_pts_attempts three_pts_pct   ft_made   ft_attempts
## Min.      : 0.00   Min.      : 0.00   Min.      : 0.00   Min.      : 0.0
## 1st Qu.: 10.00   1st Qu.: 25.00   1st Qu.: 6.00   1st Qu.: 8.0
## Median : 53.00   Median : 34.35   Median : 22.00   Median : 29.5
## Mean      : 75.78   Mean      : 29.85   Mean      : 37.22   Mean      : 47.8
## 3rd Qu.:123.00   3rd Qu.: 39.02   3rd Qu.: 49.00   3rd Qu.: 64.5
## Max.      :411.00   Max.      :100.00   Max.      :298.00   Max.      :363.0
##   ft_pct
## Min.      : 0.00
## 1st Qu.: 63.20
## Median : 76.05
## Mean      : 69.76
## 3rd Qu.: 84.62
## Max.      :100.00
```

I have selected the 3 point shooting leaders based on percentage among those players who have made over 100 3 point attempts so that players who have played a game or two and who have made 8 out of their 10 3 point attempts don't show up as the leaders. You will notice that I also do this for field goal and free throw attempts.

```
nba_3p_stats <- nba_player_stats_df %>% filter(three_pts_attempts >= 100) %>% select('Player', 'games_pl
(nba_3p_leaders <- head(nba_3p_stats[with(nba_3p_stats, order(-three_pts_pct)),],5))
```

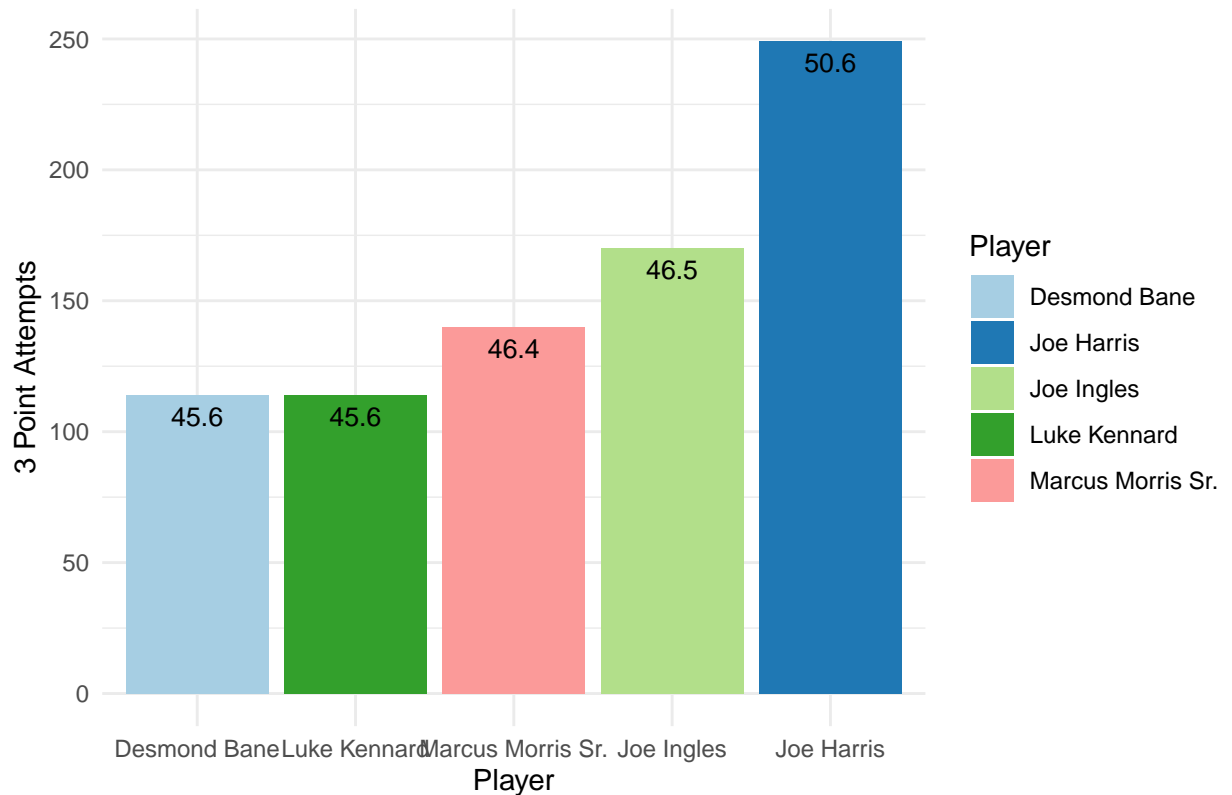
```
##           Player games_played three_pts_made three_pts_attempts
## 47      Joe Harris           37           126           249
## 94      Joe Ingles           32            79           170
## 104  Marcus Morris Sr.        28            65           140
## 123    Desmond Bane           29            52           114
## 136    Luke Kennard           34            52           114
##   three_pts_pct
## 47           50.6
## 94           46.5
## 104          46.4
## 123          45.6
## 136          45.6
```

```
nba_3p_leaders %>%
  ggplot(aes(x=reorder(Player,three_pts_attempts),y=three_pts_attempts,fill=Player)) +
  geom_bar(stat = 'identity',position=position_dodge()) +
  geom_text(aes(label=three_pts_pct), vjust=1.6, color="black",
            position = position_dodge(0.9), size=3.5) +
```

```
labs(y = ("3 Point Attempts"), x = ("Player"),
     title = ("Percentage of 3 Pointers Made by 3 Point Attempts Taken by Player") ) +
scale_fill_brewer(palette="Paired") +
theme_minimal()
```

The first statistic that I cover is the 3 point shot and shown below are the top 5 3 point shooters in the league as of 3/9/2021 through 37 games. It's hard to imagine that Joe Harris has already 249 3 point attempts and he leads the league with making just over half. Luke Kennard and Desmond Bane seem to be more efficient 3 point shooters at almost 46% only attempting 114 3

Percentage of 3 Pointers Made by 3 Point Attempts Taken by Player



pointers.

Next are the field goal shooting leaders based on percentage among those players who have made over 175 field goal attempts.

```
nba_fg_stats <- nba_player_stats_df %>% filter(fg_attempts >= 175) %>% select('Player', 'games_played', 'fg_made', 'fg_attempts', 'fg_pct')
(nba_fg_leaders <- head(nba_fg_stats[with(nba_fg_stats, order(-fg_pct)),], 5))
```

##	Player	games_played	fg_made	fg_attempts	fg_pct
## 76	Jarrett Allen	35	172	257	66.9
## 155	Ivica Zubac	38	120	182	65.9
## 97	Richaun Holmes	32	179	275	65.1
## 62	Rudy Gobert	36	196	306	64.1
## 61	Montrezl Harrell	37	209	334	62.6

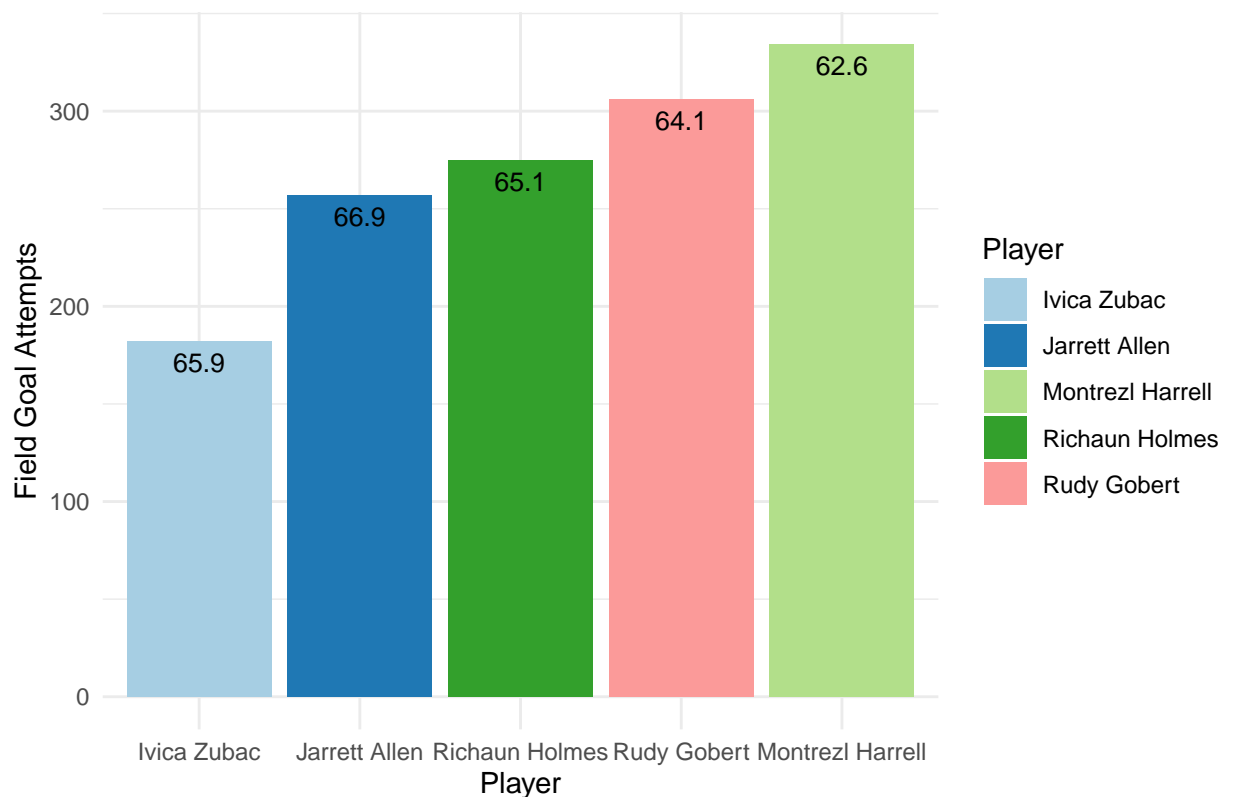
Field goals are the shots that count as 2 points. These are your high percentage shots and can be very exciting when players drive in for reverse layups, spinning dunks or “in your face” stuffs into the rim. Personally, I love to see high percentage shots off the backboard like the

ones Tim Duncan was so good at. I miss that guy. Below are the top 5 in the NBA for field goals made. I was surprised that Jarrett Allen leads with 66.9 %. For some reason I thought that these would be higher percentages.

```
nba_fg_leaders %>%
  ggplot(aes(x=reorder(Player,fg_attempts),y=fg_attempts,fill=Player)) +
  geom_bar(stat = 'identity',position=position_dodge()) +
  geom_text(aes(label=fg_pct), vjust=1.6, color="black",
            position = position_dodge(0.9), size=3.5) +

  labs(y = ("Field Goal Attempts"),x = ("Player"),
       title = ("Percentage of Field Goals Made by Field Goal Attempts Taken by Player")) +
  scale_fill_brewer(palette="Paired") +
  theme_minimal()
```

The most field goal attempts by a player is 386 by Bradley Beal and he has a 47.6 field goal per-
Percentage of Field Goals Made by Field Goal Attempts Taken by Player



centage.

Lastly, the free throw leaders are based on percentage among those players who have made over 200 free throw attempts.

```
nba_ft_stats <- nba_player_stats_df %>% filter(ft_attempts >= 200) %>% select('Player','games_played','f
(nba_ft_leaders <- head(nba_ft_stats[with(nba_ft_stats, order(-ft_pct)),],5))
```

```
##           Player games_played ft_made ft_attempts ft_pct
```

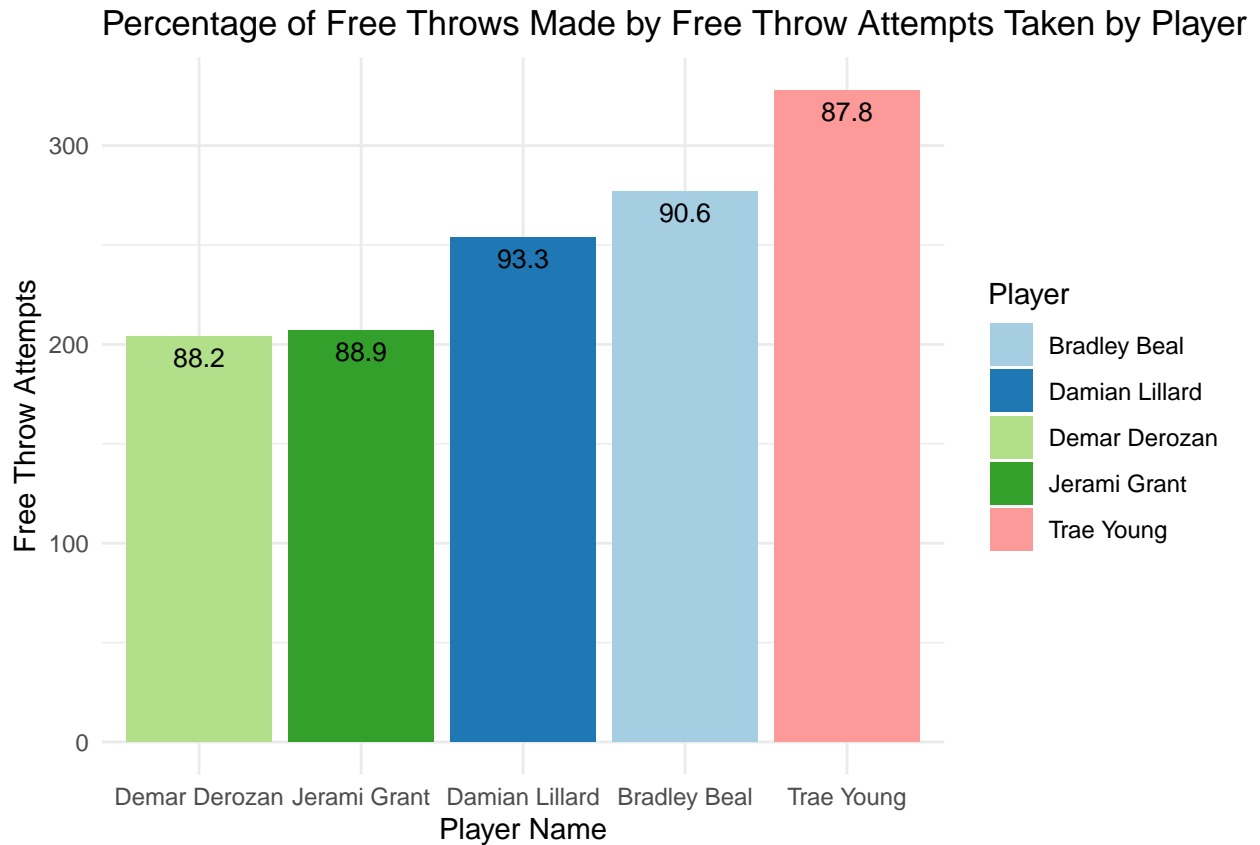
## 3	Damian Lillard	34	237	254	93.3
## 1	Bradley Beal	33	251	277	90.6
## 12	Jerami Grant	33	184	207	88.9
## 14	Demar Derozan	30	180	204	88.2
## 7	Trae Young	34	288	328	87.8

Last but certainly not least and probably the least exciting of all shots is the free throw. For obvious reasons this is the highest percentage shot in basketball. The shot is not defended and you have clear sight to the basket and you are standing still thus the reason it is called a free throw. Damian Lillard leads the league at 93.3% hitting a little over 9 out 10 free throws.

```
nba_ft_leaders %>%
  ggplot(aes(x=reorder(Player,ft_attempts),y=ft_attempts,fill=Player)) +
  geom_bar(stat = 'identity',position=position_dodge()) +
  geom_text(aes(label=ft_pct), vjust=1.6, color="black",
            position = position_dodge(0.9), size=3.5) +

  labs(y = ("Free Throw Attempts"),x = ("Player Name"),
       title = ("Percentage of Free Throws Made by Free Throw Attempts Taken by Player")) +
  scale_fill_brewer(palette="Paired") +
  theme_minimal()
```

As you can see from the percentages that most players are highly skilled free throw shooters as they should be. However, some players are not so skilled. The lowest percentage from the free throw line is 63.2%. We all know who Shaq is. He was infamous for his lack of skill at the free throw line. A good defense against his teams was the “hack-a-shaq” defense. Putting him at the line would slow his team down and would most times cost them points with his misses.



Conclusion

In conclusion, the question is this, with 10-15 seconds left in a game and your team is down by 1 point, what strategy do you employ? Do you go for the win with a 3 point attempt? A 2 point shot from the perimeter for the win or do you put the ball in your best free throw shooter's hands in the hopes that he is fouled as he drives to the basket for two points. If I was Greg Poppovich I would go with percentages and put the ball in my best free throw shooter's hands. If he does not get fouled he has a good chance of making the layup or easy shot. If he does get fouled, he has an even better chance of making the free throws to either tie or win the game.