NBA Trends: Final Report

Arman Madani, Eric Zhao, Kuang Li

December 12, 2015

Introduction:

For our project, we wanted to observe trends in the National Basketball Assocation (NBA). Specifically we wanted to observe the movement toward "Small Ball." The Small Ball (SB) style of play focuses on versatility, speed, and, most importantly 3-point shooting in favor of physical strength, height, and low-post dominance. Our main question was: have successful teams in the past couple decades adopted this style of play? Or is the hype around SB merely perceived?

We believe that the term "Small Ball" may be a misnomer for an actual NBA trend wherein teams are simply taking more 3-point attempts. Thus, while height, weight, and even field goal percentage may not necessarily be decreasing, there is an increase in three-point field goal attempts.

Data Collection:

We began by searching the internet for statistics on NBA teams and champions. We first found successful (championship) teams from 1995 - 2015. For these successful teams, we were primarily looking for data tables which included: 3-point field goals attempted, 3-point field goal percentage, total field goal percentage, and total field goal percentage. Additionally, it was important to obtain data on heights and weights of players on championship teams in order to see if there is any change in the average heights and weights of starting lineups.

Furthermore, we felt it was necessary to look at NBA Draft picks and individual players as case studies. Draft data is a good indicator of types of players that teams believe they need to improve or build around; thus it would follow that if the physical attributes of particular players moved closer to SB style players over time, teams were beginning to adopt SB. We used case studies to further illustrate the physical attributes associated with adept 3-point shooters as opposed to dominant low-post players.

To collect data efficiently, we wrote several functions to collect data from internet sources. Using for loops and the paste0() function, we were able to collect data without using repetitive commands for a majority of our raw data. For example:

```
for (i in 1995:2015) {
  url_championship_stats <-
    paste0('http://www.basketball-reference.com/leagues/NBA_', i, '.html')
  download.file(url_championship_stats, paste0('rawdata/championship_stats', i, '.html'))
}
```

This for loop was used to collect data on championship statistics from 1995 - 2015. Many of the functions to collect raw data follow a similar format.

Data Cleaning and Preparation:

All of our raw data was initially in a .html format and unusable. We spent much of our time cleaning and preparing the data in order to analyze it. Some of the cleaning and preparation was simple because the .html table was straight forward. For example, our table for NBA championship teams merely included columns for year, champion, result, and opponent. We used basic functions and one lapply() function in order to prepare this specific data set before saving it as a .csv file:

```
library(XML)
nba_champions <- readHTMLTable('rawdata/Champions.html')
nba_champions <- as.data.frame(nba_champions[[2]])
nba_champions[] <- lapply(nba_champions, as.character)
colnames(nba_champions) <- c('Year', 'Champion', 'Score', 'Opponent')
nba_champions <- nba_champions[-1, ]
write.csv(nba_champions, 'data/Champions.csv')
```

Other data sets were not as straight forward so we had to use for loops, regular expressions, and subsetting to work across multiple .html files. For example, in order to obtain the appropriate data for championship teams we used the following set of functions:

```
champion_stats <- c()
nba_champions$Champion[nba_champions$Year==1995]
for (i in 1:21) {
  file.path <- paste0('data/leagues_NBA_', 1994 + i, '_team.csv')
  files <- read.csv(file.path)
  champion_stats[[i]] <-
    files[c("Team", "FGA", "FGP", "X3PA", "X3PP", "years")]
}
for (i in 1:21){
  years <- 1994 + i
  champion <- nba_champions$Champion[nba_champions$Year==years]
  year_data <- champion_stats[[i]]
  year_data$Team <- gsub('\\*', '', as.character(year_data$Team))
  champion_stats[[i]] <-
    year_data[year_data$Team == champion | year_data$Team == "Average", ]
}
all_champs <- do.call(rbind, champion_stats)
all_champs <- all_champs[-seq(from = 2, to = nrow(all_champs), by = 2L), ]
write.csv(all_champs, file = "data/champ_data.csv")
```

Within our cleaning and preparation script, we also added columns to data sets in order to create plots later on. For example, in the roster.csv data set, we included average heights and weights of championship lineups (see below).

```
for (i in 1995:2015) {
  ch <- readHTMLTable(paste0('rawdata/champion_stats', i, '.html'))
  ch <- as.data.frame(ch[[1]])
  ch["Ht"] <- unlist(lapply(ch["Ht"][[1]], to_inches))
  ch["Year"] = i
  ch["Average Height"] = mean(ch$Ht, na.rm = TRUE)
  ch["Average Weight"] = mean(as.numeric(levels(ch$Wt)), na.rm = TRUE)
  ch <- ch[c("Player", "Pos", "Ht", "Wt", "Average Height", "Average Weight", "Year")]
  champ <- rbind(champ, ch)
}
write.csv(champ, file = "data/roster.csv")
```

Data Analysis and Results:

When it came to analyzing the data, we focused on shooting and field goals which, as aforementioned, is imperative to SB styles of play. We wanted to look at the attempted field goals from year to year, the field goal percentages from year to year, and the percentage of shots that were three pointers (See Below).

```r
# Vector of total Field Goals Attempted per year
vec_fga <- all_champs$FGA

# Descriptive statistics
mean(vec_fga)
```

```
[1] 6440.857
```

```r
sd(vec_fga)
```

```
[1] 727.9513
```

```r
all_champs$Team[which.max(all_champs$FGA)]
```

```
[1] Golden State Warriors
9 Levels: Boston Celtics Chicago Bulls ... San Antonio Spurs
```

```r
all_champs$Team[which.min(all_champs$FGA)]
```

```
[1] San Antonio Spurs
9 Levels: Boston Celtics Chicago Bulls ... San Antonio Spurs
```

```r
# Vector of Field Goal % per year
vec_fgp <- all_champs$FGP * 100

# Descriptive statistics
mean(vec_fgp)
```

```
[1] 46.83333
```

```r
sd(vec_fgp)
```

```
[1] 1.376348
```

```r
all_champs$Team[which.max(all_champs$FGP)]
```

```
[1] Miami Heat
9 Levels: Boston Celtics Chicago Bulls ... San Antonio Spurs
```

```r
all_champs$Team[which.min(all_champs$FGP)]
```

```
[1] Detroit Pistons
9 Levels: Boston Celtics Chicago Bulls ... San Antonio Spurs
```

```r
# Change in the total Field Goals Attempted Year-over-Year by Successful Teams
change_in_fga <- function(x = vec_fga) {
  output <- c()
  for (i in 1:(length(x) - 1)) {
    delta_fga <- x[i + 1] - x[i]
    output <- append(output, delta_fga)
  }
  return(output)
}


# Change in the Field Goal % Year-over-Year by Successful Teams
change_in_fgp <- function(x = vec_fgp) {
  output <- c()
  for (i in 1:(length(x) - 1)) {
    delta_fgp <- x[i + 1] - x[i]
    output <- append(output, delta_fgp)
  }
  return(output)
}

# Average change in Field Goal Percentage over 20 years
mean(change_in_fgp())
```

```
[1] -0.01
```

```r
# Vector of 3-Pt FGA per year
vec_3pfga <- all_champs$X3PA

# Descriptive statistics
mean(vec_3pfga)
```

```
[1] 1410.048
```

```r
sd(vec_3pfga)
```

```
[1] 371.9186
```

```r
# Vector of 3-Pt FG % per year
vec_3pfgp <- all_champs$X3PP * 100

# Descriptive statistics
mean(vec_3pfgp)
```

```
[1] 36.23333
```

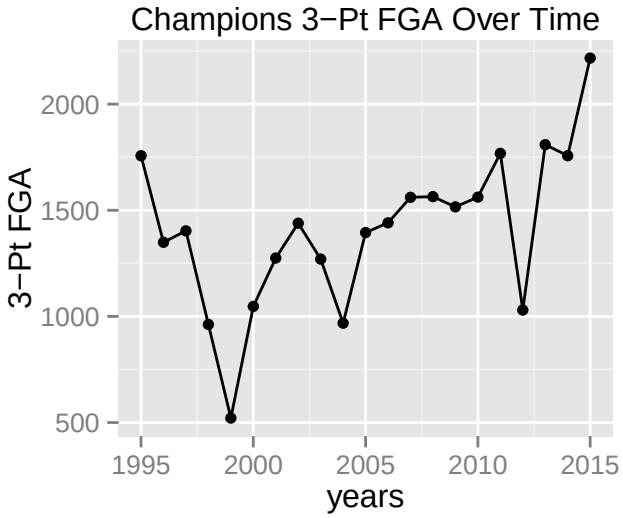
```r
sd(vec_3pfgp)
```

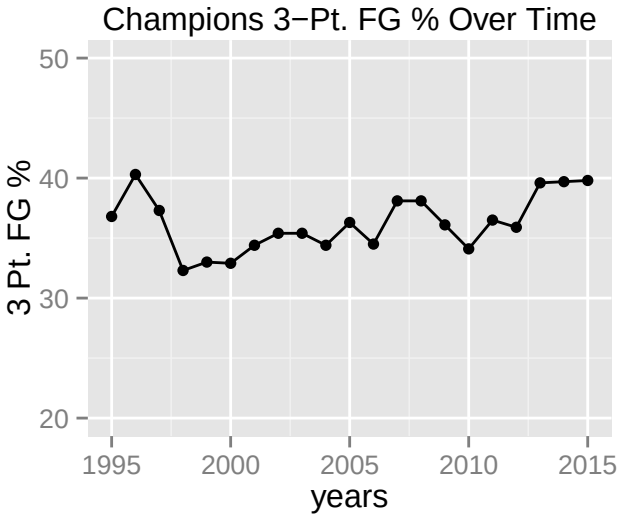
```
[1] 2.386909
```

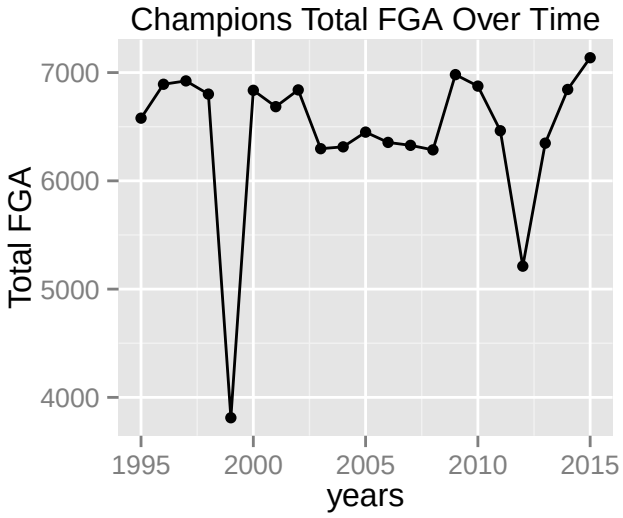
```
# 3-Pt Attempts as a Percentage of Total Field Goal Attempts (Year to Year)
vec_3pfga/vec_fga * 100
```

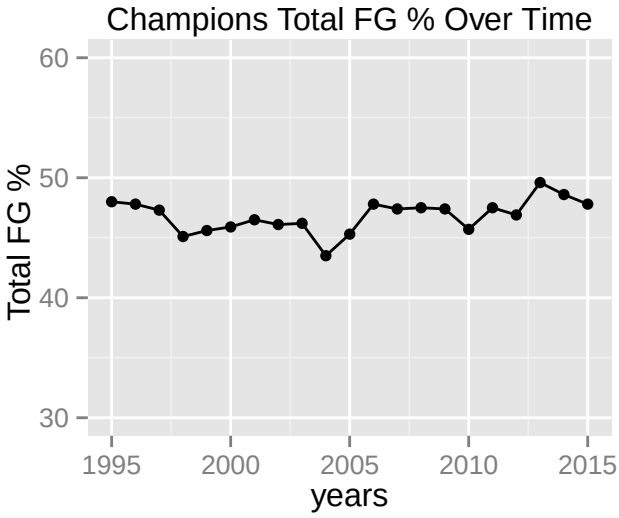
```
 [1] 26.70619 19.57342 20.26578 14.14498 13.66737 15.31597 19.07255
 [8] 21.03801 20.16833 15.33101 21.62791 22.67506 24.66814 24.88069
[15] 21.71609 22.72000 27.35572 19.76209 28.49716 25.67212 31.06347
```

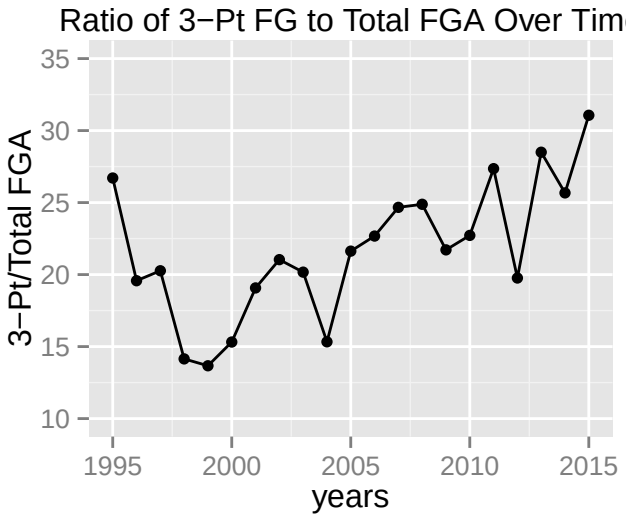
As you can see, there appears to be little to no change in 3-point field goal percentage or total field goal percentage, but 3-point attempts increased dramatically over the last 3 years. Additionally, three point attempts appear to decrease dramatically in the late 1990s and from 2008 - 2010. We believe that this is due to teams with low-post emphasis which do not shoot as many 3-pointers but have higher overall field goal percentages. However, we found that average heights and weights of players do not vary from year to year. Below, you will find visualizations which illustrate these points.

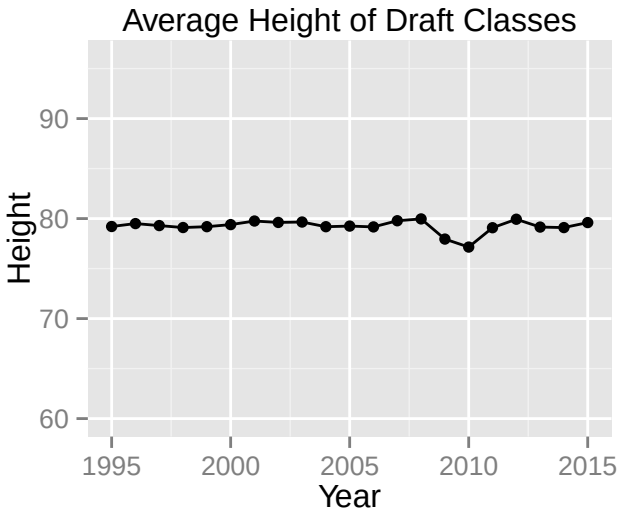


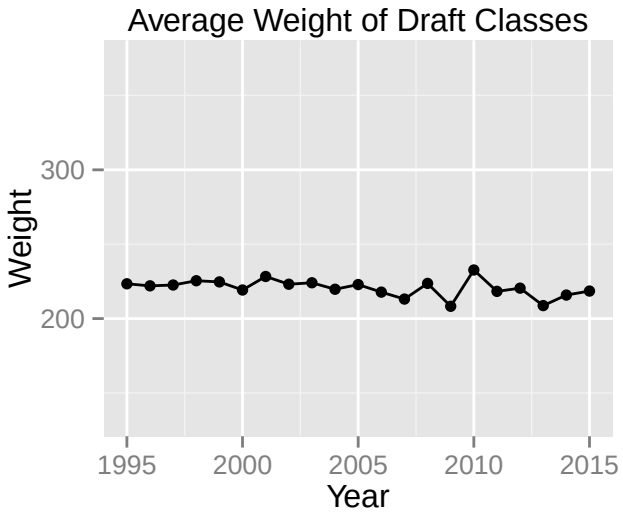


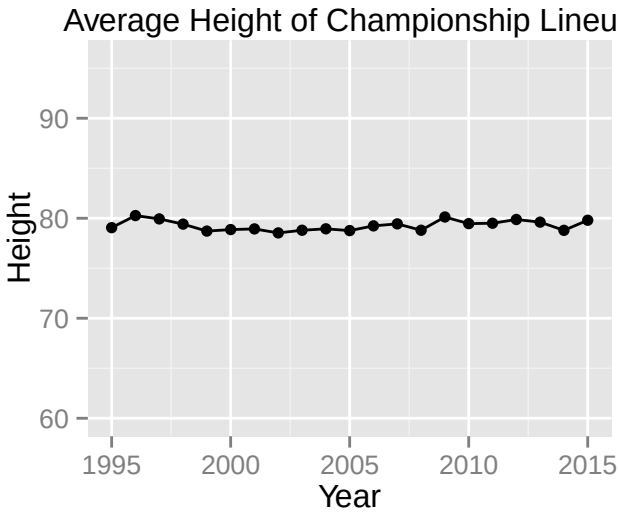


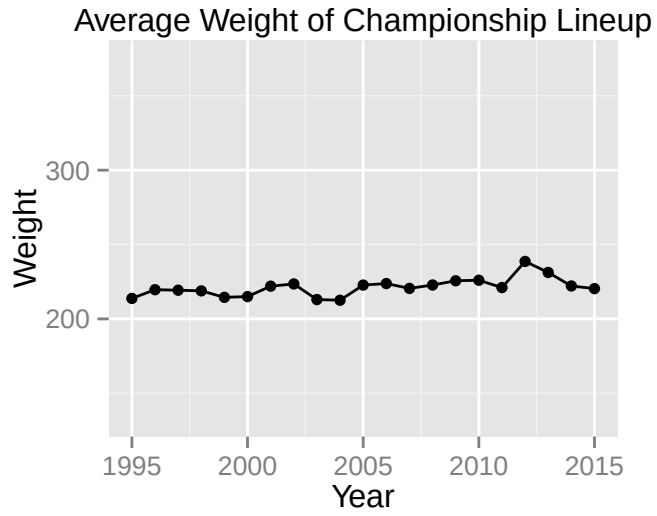



Ratio of 3−Pt FG to Total FGA Over Time
3−Pt/Total FGA
35
30
25
20
15
10
1995
2000
2005
2010
2015
years


Average Height of Draft Classes
Height
90
80
70
60
1995
2000
2005
2010
2015
Year


Average Weight of Draft Classes
Weight
300
200
1995
2000
2005
2010
2015
Year


Average Height of Championship Lineup
Height
90
80
70
60
1995
2000
2005
2010
2015
Year



The average heights and weights of both draft picks and championship lineups may not change but the individual heights and weights of key players may change.

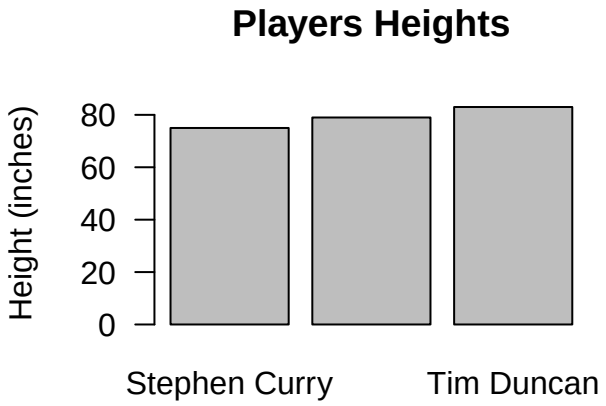
Case Study of Stephen Curry, Draymond Green, and Tim Duncan:

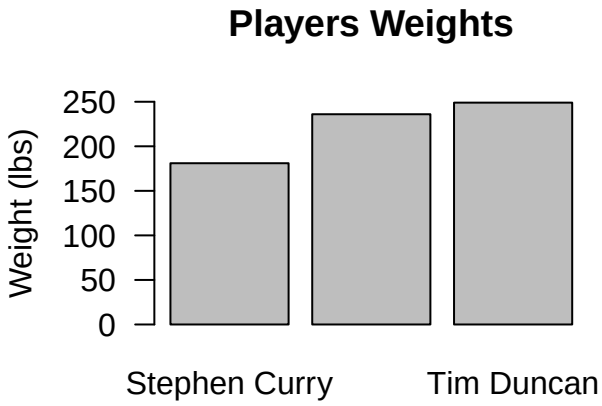
We decided to study individual players to observe specific physical attributes and shot selection. We chose Stephen Curry because he is the embodiment of the SB trend, he is not particularly tall or strong but is a 3-point shooting specialist. Draymond Green is also an example of a successful SB player, he is a power forward who is undersized for his position but shoots 3-pointers but also drives to the basketball and plays in the low-post. Tim Duncan is an oversized power forward who rarely if ever shoots 3-pointers and operates almost exclusively near the basket. Below you will find a data frame of the three players with their heights, weights, 3-point attempts, and 3-point percentage accompanied by barplots illustrating these points.

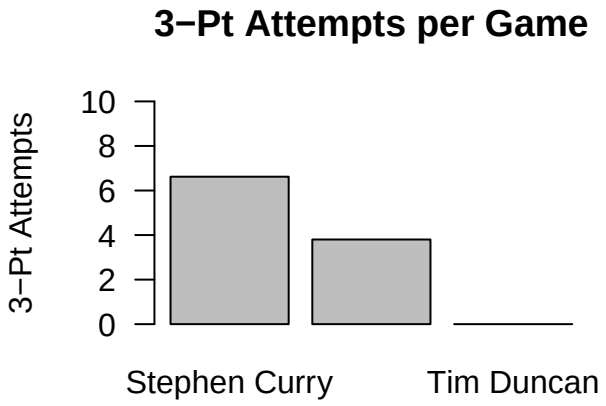
Data Frame:

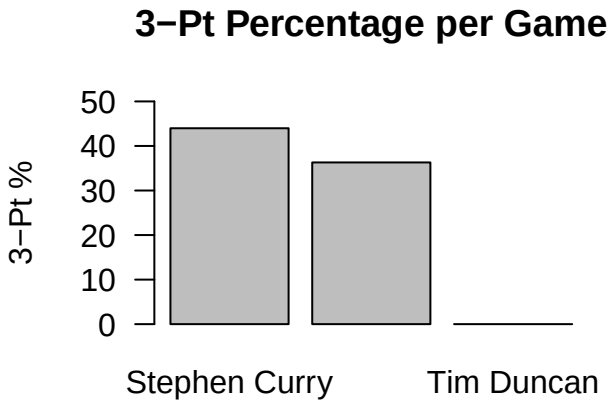
```
heights <- c(steph_curry_height, draymond_green_height, tim_duncan_height)
weights <- c(steph_curry_weight, draymond_green_weight, tim_duncan_weight)
X3PA <- c(steph_curry_3PA, draymond_green_3PA, tim_duncan_3PA)
X3PP <- c(steph_curry_3PP, draymond_green_3PP, tim_duncan_3PP)
cs_players <- c("Stephen Curry", "Draymond Green", "Tim Duncan")
case_study_df <- data.frame(cs_players, X3PA, X3PP, heights, weights)
```

Visualization:









Conclusion:

What we found through the plots and analysis was that successful teams are shooting more three pointers but their percentages are not increasing. There's also no trend toward smaller players as "small ball" would imply because the physical attributes of draft picks and championship rosters are staying relatively constant. So calling this trend "small ball" is a misnomer, teams aren't playing smaller players, they're shooting more threes with similar percentages of success which is a more efficient strategy (intuitively, expected point per shot increases).