Master

Mitch Oshima

April 13, 2016

This is walkthrough of using my shot data visualization tools. The first step in this process is to download the list of active players from nba.com. To do this we must install and load the following packages

```
require(rjson)
```

```
## Loading required package: rjson
```

```
require(ggplot2)
```

```
## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 3.1.3
```

```
require(nlme)
```

```
## Loading required package: nlme

## Warning: package 'nlme' was built under R version 3.1.3
```

Next, we'll run the following code to scrape the list of players from nba.com using a url to their json formatted data. This code can all be found in the shotDataScraper r script.

```
idURL <- paste("http://stats.nba.com/stats/commonallplayers?",
               "IsOnlyCurrentSeason=1&LeagueID=00&Season=2015-16",
               sep = "")

idData <- fromJSON(file = idURL, method="C")
idDf<- data.frame(
  matrix(unlist(idData$resultSets[[1]][[3]]), ncol = 13,byrow = TRUE),
  stringsAsFactors = FALSE)
colnames(idDf)<- idData$resultSets[[1]][[2]]
```

We can also use this data to get a list of all teams in the NBA.

```
teamIdDf <- unique(
  subset(idDf,TEAM_ID>0, select = TEAM_ID:TEAM_CODE)
)
```

To continue on with the visualization the two previous code chunks must be run. Now, we'll use the player id data now to create the first data scraping function. This function will be used to gather the shot location data from NBA.com. This function is the backbone of the rest of the project. Most of the subsequent functions rely on getShotData.

```
getShotData <- function(playerName){

  ndx <- which(idDf$DISPLAY_FIRST_LAST %in% playerName)
  player <- idDf[ndx,1]

  shotURL <- paste("http://stats.nba.com/stats/shotchartdetail?CFID=&CFPARAMS",
                   "=&ContextFilter=&ContextMeasure=FGA&DateFrom=&DateTo=&",
                   "GameID=&GameSegment=&LastNGames=0&",
                   "LeagueID=00&Location=&Month=0&OpponentTeamID=0&Outcome=&",
                   "Period=0&Position=&RookieYear=&SeasonSegment=&",
                   "SeasonType=Regular+Season&TeamID=0&VsConference=&",
                   "VsDivision=&PlayerID=",player,
                   "&Season=2015-16",
                   sep = "")
  shotData <- fromJSON(file = shotURL, method="C")
  shotDataf <- data.frame(
    matrix(unlist(shotData$resultSets[[1]][[3]]),
      ncol=21, byrow = TRUE)
  )
  #set column names
  colnames(shotDataf) <- shotData$resultSets[[1]][[2]]
  #convert coordinates to numerics
  shotDataf$LOC_X <- as.numeric(
    as.character(shotDataf$LOC_X)
  )
  shotDataf$LOC_Y <- as.numeric(
    as.character(shotDataf$LOC_Y)
  )
  shotDataf$SHOT_DISTANCE <- as.numeric(
    as.character(shotDataf$SHOT_DISTANCE)
  )
  return(shotDataf)
}
```

We'll also build a function that will look up and return the player id data for every player on a given team.

```
getRoster <- function(teamAbbr){
  tmNdx <- which(teamIdDf$TEAM_ABBREVIATION %in% teamAbbr)
  teamId <- teamIdDf[tmNdx,1]
  plNdx <- which(idDf$TEAM_ID %in% teamId)
  roster <- idDf[plNdx,3]
  return(roster)
}
```

Using the getRoster and getShotData functions we can now get the the shot data for every player on the team. This function will be useful down the road for deeper team analysis.

```
getTeamShooting <- function(team){
  teamShotData <- NULL
  ros <- getRoster(team)
  for (i in 1:length(ros)){
    playerShot <- getShotData(ros[i])
    teamShotData <- rbind(teamShotData, playerShot)
```

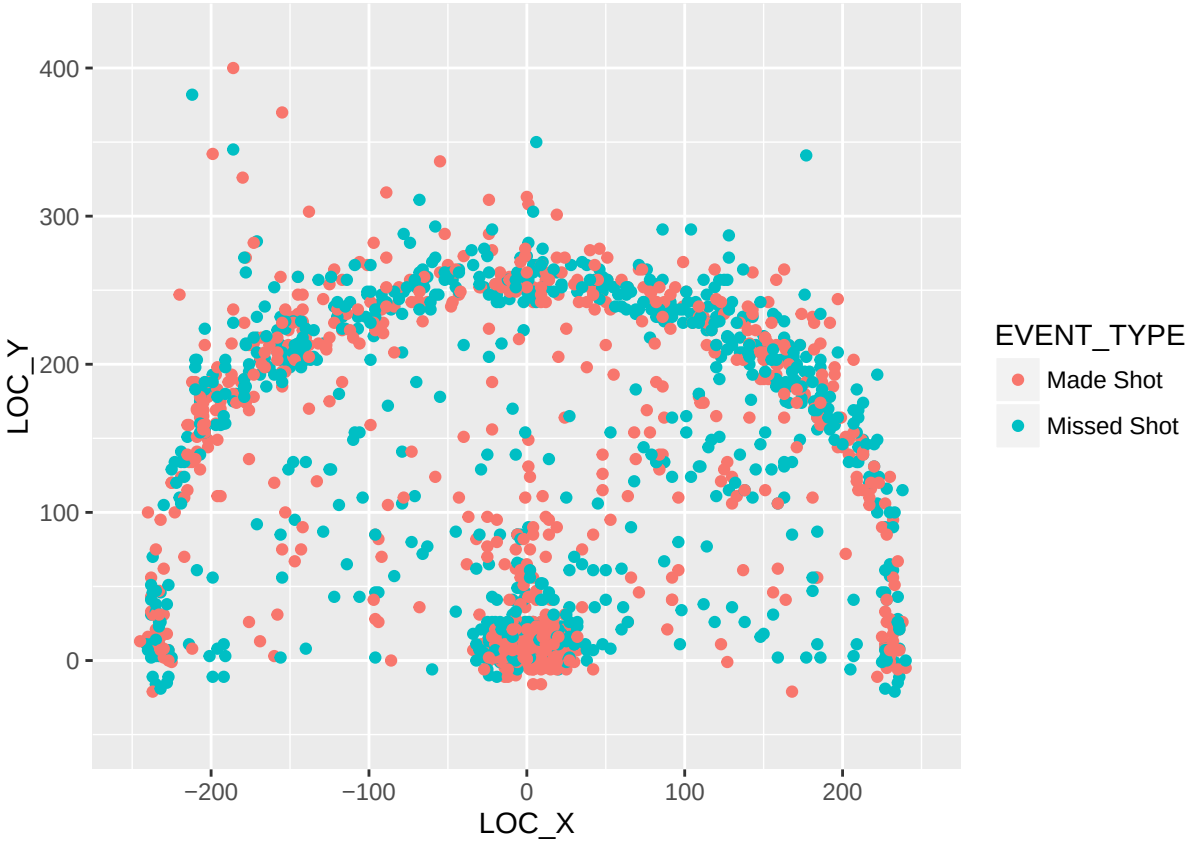
```
  }
  return(teamShotData)
}
```

Now that we have the tools to collect the shot location data we can begin to visualize it. First we can build a very basic shot chart. This shot chart simply plots the shot attempts and color codes the points based on whether the shot was a make or a miss.

```
basicShotChart <- function(playerName){
  df <- getShotData(playerName)
  shotChart <- ggplot(df,aes(x=LOC_X, y=LOC_Y)) +
    geom_point(aes(colour = EVENT_TYPE))+
    xlim(-250,250)+
    ylim(-50,420)
  return(shotChart)
}
```

Let's see what this looks like for Stephen Curry

```
## Warning: Removed 11 rows containing missing values (geom_point).
```



Not bad, but with a little more work we can make it even better. To do this we will make a model to predict whether the shot will be a make or a miss based on the (x,y) location of the attempt.

```
shotModelFit <- function(player){
  df <- getShotData(player)
  mod <- glm(SHOT_MADE_FLAG ~ LOC_X + LOC_Y,
             data = df,
             family = binomial)
  prdct <- df[,c("LOC_X","LOC_Y")]
  df$PROB <- predict(mod, newdata = prdct)
  return(df)
}
```

We will now plot the shot data and color the points based on the probability of a make. We will also add a density layer to the plot to show which locations on the floor have the highest concentration of shot attempts.

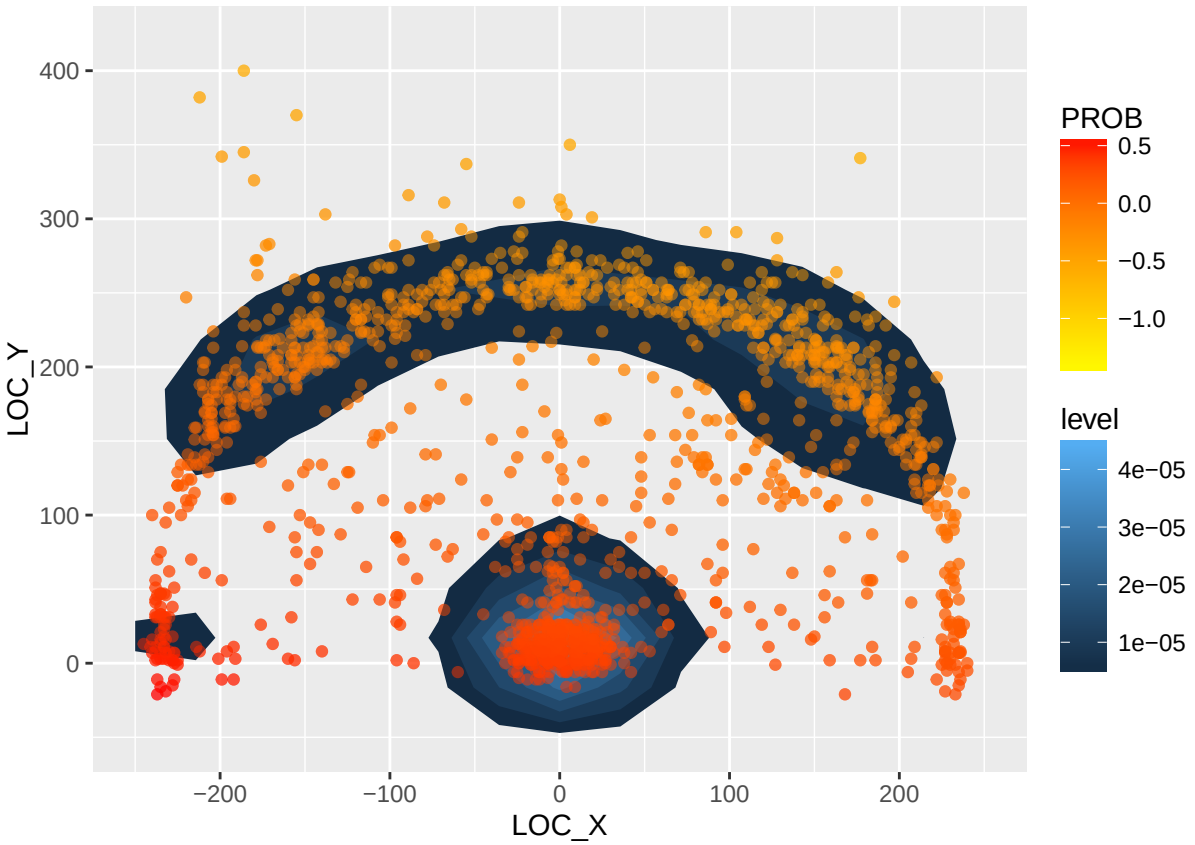
```
densityShotChart <- function(player){
  df <- shotModelFit(player)
  shotChart <- ggplot(data = df, aes(x=LOC_X, y=LOC_Y, z=PROB)) +
      xlim(-250,250)+
      ylim(-50,420)+
      geom_point(aes(color = PROB), alpha = 1/2 ) +
      scale_color_gradient(low="yellow",high="red")+
      stat_density2d(geom = "polygon", n=15, aes(fill=..level..))+
      geom_point(aes(color = PROB), alpha = 1/2)
  return(shotChart)
}
```

Now let's see what Stephen Curry's density shot chart looks like.

```
## Warning: Removed 11 rows containing non-finite values (stat_density2d).

## Warning: Removed 11 rows containing missing values (geom_point).

## Warning: Removed 11 rows containing missing values (geom_point).
```



Pretty cool! The code for the shot charts can also be found in the shotCharts r script.