

Analyzing and Visualizing Jeremy Lin's stats

Jenny Huang

2017.12.02

1.Introduction

We used basic data analysis and graphical tools in class to analyze and visualize the NBA data, so this post aims to delve deeper into sports statistics and visualization of it in more game-relevant ways.

For example, we want to be able to see how the data extracted applies in context of different areas of the basketball court

By the end of the post, we want to be able to diagrammatically present advanced NBA data in a more tangible and complex ways and analyze the visualized data accordingly.

Also, we will use packages such as `ggplot2`, `jpeg`, `grid`, and `Rcurl`.

By using these methods, we will be able to visualize and strategize in-game more easily by putting data on a basketball court.

It will also help see the most suitable attacking options as well as defensive options for players. For example, If the data shows that the opponent's player specializes in 3 pointers, a coach may ask for a tighter defense around the 3 point line against the particular player so that he is unable to attempt those 3 pointers frequently.

2.Preparation

First, we need to load `ggplot2` and `dplyr` to help us with data analysis and visualization.

```
library(ggplot2)
library(dplyr)
```

```
##
## 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
```

Since the database for all point guards is quite extensive and requires us to compile all players' data together, I'm going to focus on one of my favorite players, Jeremy Lin, in this post.

3.Data processing and presenting

I downloaded Jeremy Lin's player stats from a website called NBAsavant.com. This csv file shows his statistics during 2016-2017 season.

3.1.Importing data

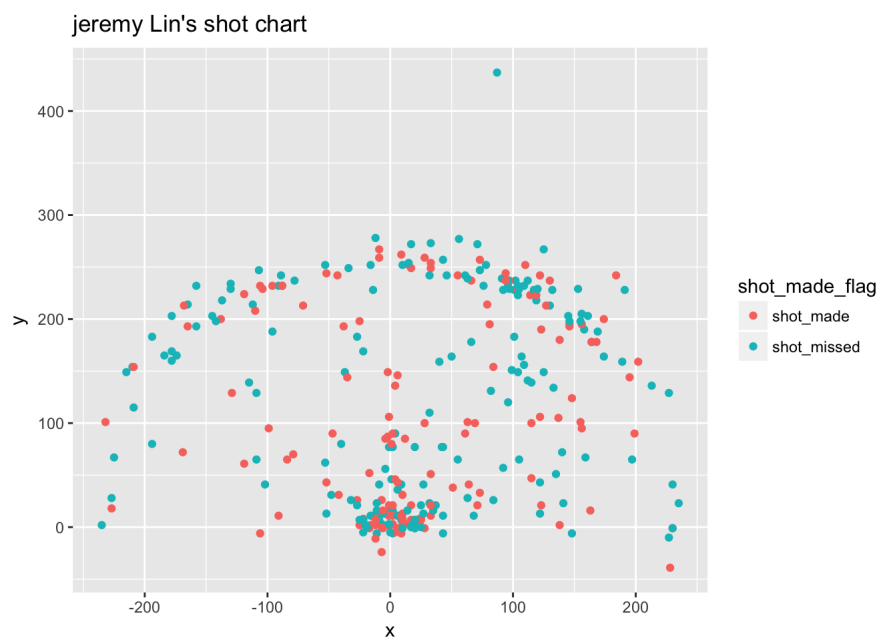
```
#importing csv file, and classifying columns
jeremy <- read.csv("nba_savant202391.csv", colClasses =
  c("name" = "character",
    "action_type" = "character",
    "shot_type" = "character",
    "shot_made_flag" = "character"
  ))
str(jeremy)
```

```
## 'data.frame': 331 obs. of 22 variables:
## $ name : chr "Jeremy Lin" "Jeremy Lin" "Jeremy Lin" "Jeremy Lin" ...
## $ team_name : Factor w/ 1 level "Brooklyn Nets": 1 1 1 1 1 1 1 1 1 1 ...
## $ game_date : Factor w/ 30 levels "2016-10-26","2016-10-28",...: 25 30 5 14 15 17 6 30 17 20 ...
## $ season : int 2016 2016 2016 2016 2016 2016 2016 2016 2016 2016 ...
## $ espn_player_id : int 4299 4299 4299 4299 4299 4299 4299 4299 4299 4299 ...
## $ team_id : int 1610612751 1610612751 1610612751 1610612751 1610612751 1610612751 1610612751 1610612751 1610612751 1610612751 ...
## $ espn_game_id : int 400900467 400900552 400899507 400900279 400899448 400900321 400899806 400900552 400900321 400900395 ...
## $ period : int 2 1 1 1 3 1 3 4 4 2 ...
## $ minutes_remaining: int 6 10 6 9 10 9 0 0 4 0 ...
## $ seconds_remaining: int 23 41 29 51 41 19 31 24 8 42 ...
## $ shot_made_flag : chr "1" "1" "1" "1" ...
## $ action_type : chr "Cutting Layup Shot" "Cutting Layup Shot" "Cutting Layup Shot" "Cutting Layup Shot" ...
## $ shot_type : chr "2PT Field Goal" "2PT Field Goal" "2PT Field Goal" "2PT Field Goal" ...
## $ shot_distance : int 1 2 0 1 2 0 6 1 4 3 ...
## $ opponent : Factor w/ 23 levels "Atlanta Hawks",...: 23 16 8 9 20 19 10 16 19 15 ...
## $ x : int 15 24 0 -12 -22 2 -52 -1 43 -27 ...
## $ y : int 3 7 2 11 8 -6 43 16 11 26 ...
## $ dribbles : int 0 0 0 0 0 0 0 0 0 0 ...
## $ touch_time : num 0 0 0 0 0 0 0 0 0 0 ...
## $ defender_name : logi NA NA NA NA NA NA ...
## $ defender_distance: num 0 0 0 0 0 0 0 0 0 0 ...
## $ shot_clock : num 0 0 0 0 0 0 0 0 0 0 ...
```

In the dataframe of Jeremy, we are also given the x and y coordinates of every shot taken. We can use this to create a plot showing the shots taken at different locations. We can also label them depending on whether the shots were made or missed. In the csv file and the `shot_made_flag` column, 1 represents a shot made and 0 represents a shot missed.

3.2.Scatterplot of shots made and missed

```
#first, converting the shot_made_flag values from 0 and 1 to made and missed
jeremy$shot_made_flag[jeremy$shot_made_flag == 1] <- "shot_made"
jeremy$shot_made_flag[jeremy$shot_made_flag == 0] <- "shot_missed"
#creating a plot using dataframe
ggplot(jeremy, aes(x = x, y = y)) + geom_point(aes(colour = shot_made_flag)) + ggtitle("jeremy Lin's shot chart")
```



In this plot, we can tell where Jeremy tends to shoot frequently, and whether he misses or makes those shoots. At a first glance, this gives us a pretty clear visualization and help us get a grasp of the player's accuracy.

From this plot, we can kinda see a shape of the basketball court. However, we want to distinguish the type of shots attempted as well as where these points lie in relation to the court structure. We can do this by overlaying an image of a basketball court on the `ggplot`, and using packages `grid`, `jpeg`, and `Rcurl` to obtain the image. We can use `annotation_custom` in `ggplot` to complete the overlay.

3.3.Plot with halfcourt image

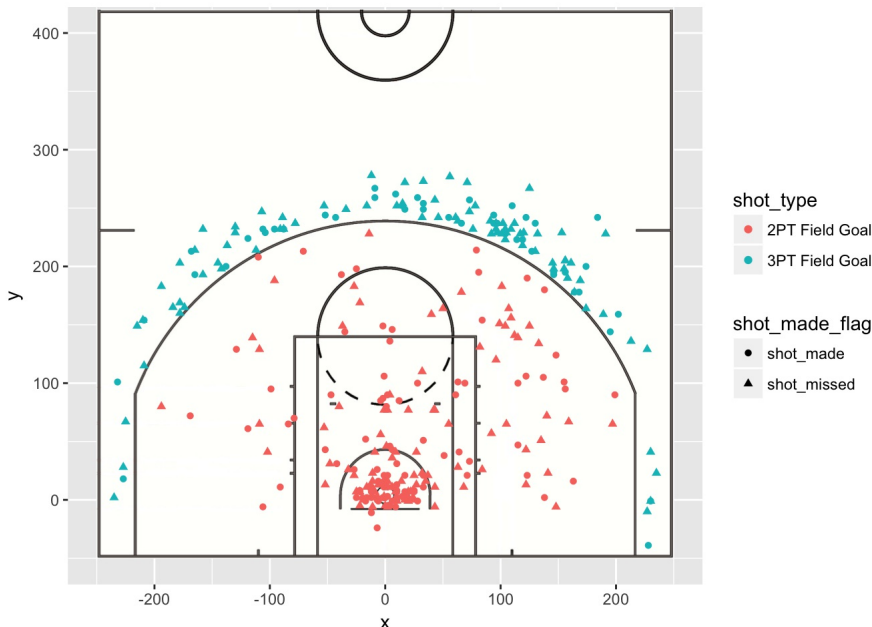
```
#loading packages
library(grid)
library(jpeg)
library(Rcurl)
```

```
## Loading required package: bitops
```

```
#downloading the halfcourt image
halfcourt.URL <- "https://thedatagame.files.wordpress.com/2016/03/nba_court.jpg"
halfcourt <- rasterGrob(readJPEG(getURLContent(halfcourt.URL)),
                        width = unit(1, "npc"), height = unit(1, "npc"))

#performing the ggplot
ggplot(jeremy, aes(x = x, y = y)) + annotation_custom(halfcourt, -250,250,-50,420) +
  geom_point(aes(colour = shot_type , shape = shot_made_flag)) +
  xlim(-250, 250) +
  ylim(-50, 400)
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```



```
#let's also save this plot
png(file = "jeremy_shotchart.png")
ggplot(jeremy, aes(x = x, y = y)) + annotation_custom(halfcourt, -250,250,-50,420) +
  geom_point(aes(colour = shot_type , shape = shot_made_flag)) +
  xlim(-250, 250) +
  ylim(-50, 400) +
  ggtitle("Jeremy Lin - Shot Chart")
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

```
dev.off()
```

```
## quartz_off_screen
##                2
```

The ratio of this plot is off and doesn't represent the ratio of an actual court, but we can fix it by using the `coord_fixed` function in `ggplot`.

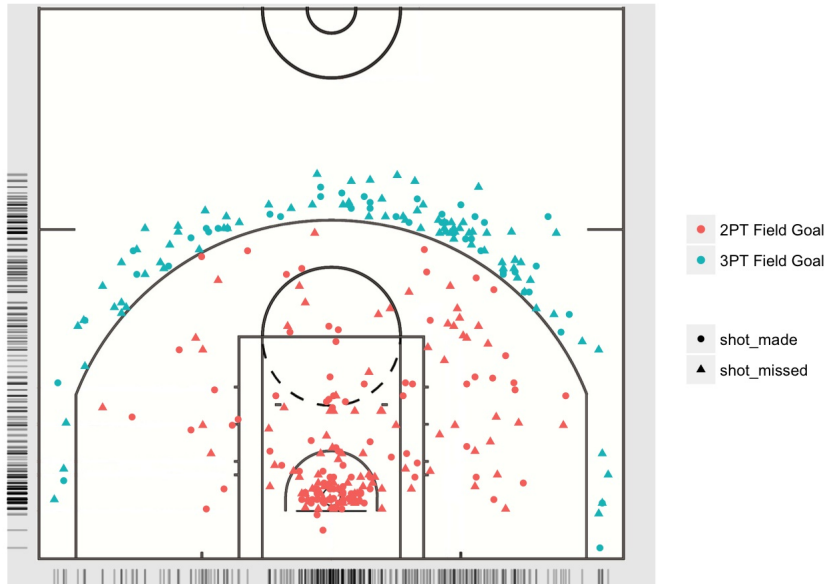
3.4. Another plot

As we know that the x axis represents the x coordinate of the shot and y axis the y coordinate of the shot, I decided to remove those axes. Also, I decided to use `geom_rug` to display the frequency of the shots in the x,y coordinates.

```
#replotting using ggplot
ggplot(jeremy, aes(x = x, y = y)) + annotation_custom(halfcourt, -250,250,-50,420) +
  geom_point(aes(colour = shot_type , shape = shot_made_flag)) +
  xlim(-250, 250) +
  ylim(-50, 400) +
  geom_rug(alpha = 0.25) +
  coord_fixed() +
  ggtitle("Jeremy Lin - Shot Chart") +
  theme(line = element_blank(),
        axis.title.x = element_blank(),
        axis.title.y = element_blank(),
        legend.title = element_blank(),
        axis.text.x = element_blank(),
        axis.text.y = element_blank()
  )
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

Jeremy Lin - Shot Chart



```
png(file = "jeremy_shotchart2.png")
ggplot(jeremy, aes(x = x, y = y)) + annotation_custom(halfcourt, -250,250,-50,420) +
  geom_point(aes(colour = shot_type , shape = shot_made_flag)) +
  xlim(-250, 250) +
  ylim(-50, 400) +
  geom_rug(alpha = 0.25) +
  coord_fixed() +
  ggtitle("Jeremy Lin - Shot Chart") +
  theme(line = element_blank(),
        axis.title.x = element_blank(),
        axis.title.y = element_blank(),
        legend.title = element_blank(),
        axis.text.x = element_blank(),
        axis.text.y = element_blank()
  )
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

```
dev.off()
```

```
## quartz_off_screen
##                2
```

Here, we can see that the plot is much cleaner and perfectly resemblant of the half court. As we observed above and here again through the `geom_rug` feature, it seems to be that Jeremy's shots are mostly outside the perimeter or drive into the basket.

4.Conclusion

We have been exposed to several graphing packages in this class, and each package has its advantages in visualizing data and giving us general ideas about the dataset. For the analysis purpose of this post where we want to analyze basketball player's stats, I think it is more straightforward to plot the data points on a chart that resembles the shape of a basketball, which gives us a very straightforward and clear picture.

In terms of the analysis itself relating to Jeremy Lin and his shooting style, it seems that he prefers to shoot on the perimeter or drive into the basket. In defense's perspective, it is important to guard him tightly on the perimeter and have someone constantly guarding near the basket to stop him from easy layups.

From offense's perspective, it is important to utilize Jeremy's ability and advantage in shooting 3 pointers.

4.1 Take-home Message

Graphical presentation usually offers us a better conceptual idea of what our dataset entails, and choosing the ideal type of representation will give us a clearer and more straightforward generalization that helps us to answer the topic we are researching on. `ggplot` comes in really handy in this case as it offers a more simplistic and realistic representation of our data. On top of that, being able to overlay an image of the basketball court also aids us in the process of analyzing a player's shooting style.

5.References

<https://cran.r-project.org/web/packages/hexbin/hexbin.pdf>

<https://www.rdocumentation.org/packages/grid/versions/3.4.1>

<https://cran.r-project.org/web/packages/Rcurl/Rcurl.pdf>

<https://thedatagame.com.au/2015/09/27/how-to-create-nba-shot-charts-in-r/>

http://nbasavant.com/shot_search.php

<https://themarkscard.com/2016/11/30/scraping-and-organizing-nba-shot-chart-data-with-python-and-the-tableau-extract-api/>

<http://r-statistics.co/ggplot2-cheatsheet.html>

<https://cran.r-project.org/web/packages/jpeg/jpeg.pdf>