stat442final project

Robin

2022-12-08

Introduction

Every NBA season, there will be a selection of the Most Valuable Player (MVP) award. It is one of the most prestigious accolades that an NBA player can receive in his career. The selection is conducted by independent media members who are not affiliated with teams nor players. Each member selects players in a weighted voting system. From 2010, public can also participate in voting the MVP. However, there are so many irrational factors such as public image and scandals that can affect the selection of a player. In my project, I will only focus on the performance of playing basketball, using 4 visualization, based on the data NBA_Players_Boxscore, to choose my MVP.

Data preparation

There are over 300 players in the league, while most of them are not eligible for being MVP, so I will filter the data first. The eligible players should be:
- starters for every game played
- average playing time per game more than 25 minutes
- scores over 23 points per game

```
nba <- read.csv("C:/Users/76506/Desktop/stat442 final/NBA_Player_Boxscore_2021-22.csv",
                stringsAsFactors=TRUE)
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
```

```
# average playing time per game more than 25 minutes
df <- nba %>%  group_by(athlete_display_name) %>%
  summarise_at(vars(min),list(avgtime = mean)) %>%
  filter(avgtime > 25)
nba <- subset(nba,athlete_display_name %in% df$athlete_display_name)
```

```
# starters for every game played
df <- nba %>% group_by(athlete_display_name) %>% filter(starter == FALSE)
nba <- subset(nba,!athlete_display_name %in% df$athlete_display_name)
# scores over 20 points per game
df <- nba %>% group_by(athlete_display_name) %>%
  summarise_at(vars(pts),list(avgpts = mean)) %>%
  filter(avgpts > 23)
nba <- subset(nba,athlete_display_name %in% df$athlete_display_name)
# qualified players
players <- unique(nba$athlete_display_name)
length(players)
```

```
## [1] 19
```

After filtering by basic requirements, there are 19 players left for later discussion.

Visualization

Viz1: Wildcard (a huxtable of players' average rebounds and assists sorted by mean points)

PPG: points per game RPG: rebounds per game APG: assists per game

```
library(huxtable)
```

```
##
## Attaching package: 'huxtable'
```

```
## The following object is masked from 'package:dplyr':
##
##     add_rownames
```

```
library(ggplot2)
```

```
##
## Attaching package: 'ggplot2'
```

```
## The following object is masked from 'package:huxtable':
##
##     theme_grey
```

```
library(knitr)
df <- nba %>% group_by(athlete_display_name) %>%
summarise('player name' = athlete_short_name[1],
'position' = athlete_position_name[1],
'PPG' = mean(pts),
'RPG' = mean(reb),
```

```
'APG' = mean(ast))%>%
arrange(-`PPG`) %>%
select(-`athlete_display_name`)
mytable <- df %>% as_huxtable() %>%
set_background_color(evens,everywhere,"grey")%>%
set_outer_borders()%>%
set_bottom_border(1,)%>%
set_bold(1,everywhere)
mytable
```

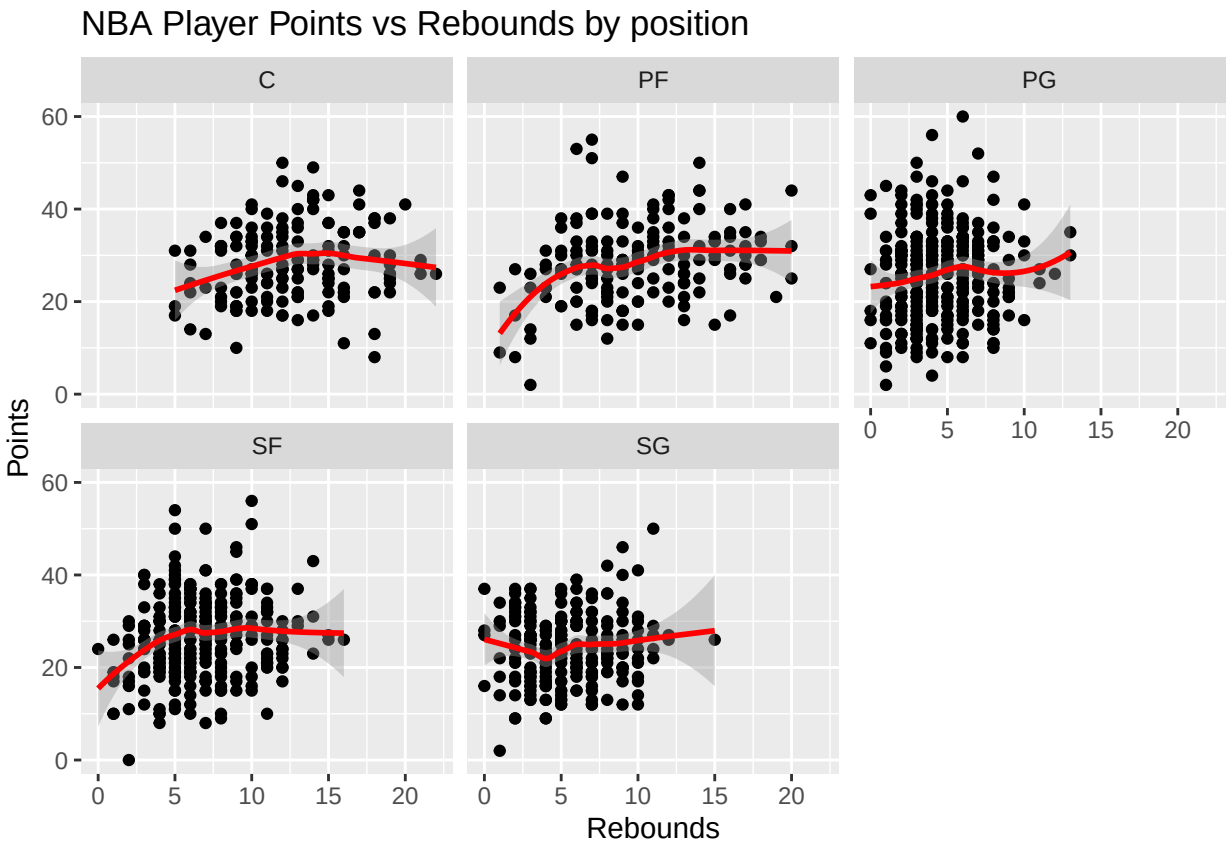
player name	position	PPG	RPG	APG
L. James	Small Forward	30.2	8.18	6.26
G. Antetokounmpo	Power Forward	30.1	12	5.94
J. Embiid	Center	29.8	11.6	3.92
K. Durant	Power Forward	29.6	7.29	6.37
T. Young	Point Guard	27.4	3.79	9.49
D. DeRozan	Small Forward	27.2	5.15	4.91
N. Jokic	Center	27.1	13.6	7.76
J. Morant	Point Guard	27.1	5.94	7.09
J. Tatum	Small Forward	26.7	7.72	4.68
K. Irving	Point Guard	26.7	4.45	5.7
D. Mitchell	Shooting Guard	25.8	4.22	5.37
S. Gilgeous-Alexander	Shooting Guard	24.5	4.96	5.91
P. George	Shooting Guard	24.4	6.87	5.68
D. Lillard	Point Guard	24	4.14	7.31
J. Brown	Shooting Guard	23.5	6.25	3.5
B. Beal	Shooting Guard	23.2	4.72	6.62
A. Davis	Power Forward	23.2	9.85	3.05
D. Fox	Point Guard	23.2	3.92	5.58
B. Ingram	Small Forward	23.1	5.85	5.62

We can find that players such as Giannis, Lebron, and Jokic have not only scored many points, their rebounds and assists are also better than most of others. Because of the basketball rules, players on different positions have different data structure. Generally, centers are responsible for getting rebounds, guards have more assists and forwards are more comprehensive. Therefore, we need to facet the players by their position.

Viz2 categorical(facet by positions)

```
mygraph <- nba %>% ggplot(aes(x=reb,y=pts)) +
geom_point() +
geom_smooth(method = "loess",
se = TRUE,col = "red") +
facet_wrap(~athlete_position_abbreviation) +
labs(title = "NBA Player Points vs Rebounds by position",
x = "Rebounds",
y = "Points")
mygraph
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



As shown in the graph, because of the basketball rules, players on different positions have different data structure. Generally, centers are responsible for getting rebounds, guards have more assists and scores, forwards are more comprehensive. Therefore, it is not wise to compare players' stats directly.

Viz3 2D (dot plots for different positions)

Let's plot three dot plots, one for centers, one for forwards, one for guards, and find the best player in each position. Let's use the first 2 or 3 players with highest PPG in each position, based on the huxtable from Viz1.

```
library(ggpubr)
```

```
##
## Attaching package: 'ggpubr'
```

```
## The following object is masked from 'package:huxtable':
##
##     font
```

```
library("patchwork")
df_forward <- subset(nba,athlete_display_name %in% c('LeBron James', 'Giannis Antetokounmpo','Kevin Dura
df_center <- subset(nba,athlete_display_name %in% c('Joel Embiid', 'Nikola Jokic'))
df_guard <- subset(nba,athlete_display_name %in% c('Trae Young', 'Ja Morant','Kyrie Irving'))
# box-dot plot
plot_center <- ggplot(df_center,
                      aes(x=athlete_short_name, y=reb)) + geom_boxplot()+
  geom_dotplot(binaxis='y', stackdir='center',binwidth = 0.5)+
  labs(x = "player name",y ="rebounds")
# violin-dot plot
plot_forward <- ggplot(df_forward, aes(x=athlete_short_name, y=pts)) +geom_violin(trim = FALSE)+
  geom_dotplot(binaxis='y', stackdir='center',binwidth = 1.5)+
  labs(x = "player name",y ="points")
# notched-box plot
plot_guard <- ggplot(df_guard, aes(x=athlete_short_name, y=ast)) + geom_boxplot(notch = TRUE)+
  geom_dotplot(binaxis='y', stackdir='center',binwidth = 0.5)+
  labs(x = "player name",y ="assists")
ggarrange(plot_center, plot_forward, plot_guard,
          labels = c("center", "forward", "guard"),
          ncol = 2, nrow = 2) +
  plot_annotation(title = "dot plots faceted by positions") &
  theme(plot.title = element_text(hjust = 0.5))
```

dot plots faceted by positions