HW7

Shero_Dawson_HW3

2024-02-22

Packages Required

```
pacman::p_load(tidyverse, hoopR, cowplot, here)
```

Data

```
shots18 <- read.csv(here("data","shots_1819.csv"))

players <- read.csv(here("data","nba_players.csv")) %>%
  select(PERSON_ID, DISPLAY_FIRST_LAST)
```

HOMEWORK

Here are the questions for your homework. You should answer these in an RMarkdown document, knit to **PDF** and submit both your **.Rmd** and **PDF** on Canvas. We're going to start by identifying Mr. Analytics for the 2018-19 NBA season.

HW Q1: First, read in the shot chart data, save it as `shots18`, and then **modify** the code below to add two variables: the percent of shots taken from 3-pt range, and then the *mean distance* on 2-pt shots. (1 pt)

Data Showing Different Shots Per Player in The 2018-19 Season

```
players18 <- shots18 %>%
  mutate(LOC_X = LOC_X/10,
         LOC_Y = LOC_Y/10,
         SHOT_DISTANCE = sqrt(LOC_X^2 + LOC_Y^2)) %>%
  mutate(pts = case_when(SHOT_TYPE == "3PT Field Goal" ~ 3,
                         TRUE ~ 2)) %>%
  group_by(PLAYER_NAME, TEAM_NAME) %>%
  summarize(FGA = n(),
            FGM = sum(SHOT_MADE_FLAG),
            FGA_2pt = sum(pts[pts==2])/2,
            FGA_3pt = sum(pts[pts==3])/3,
```

```
            FGM_2pt = sum(pts[pts==2]*SHOT_MADE_FLAG[pts==2])/2,
            FGM_3pt = sum(pts[pts==3]*SHOT_MADE_FLAG[pts==3])/3,
            FGP_2pt = ((FGM_2pt/FGA_2pt) * 100),
            FGP_3pt = ((FGM_3pt/FGA_3pt) * 100),
            MEAN_DISTANCE_2pt = mean(SHOT_DISTANCE <= 24), .groups = "drop")
```

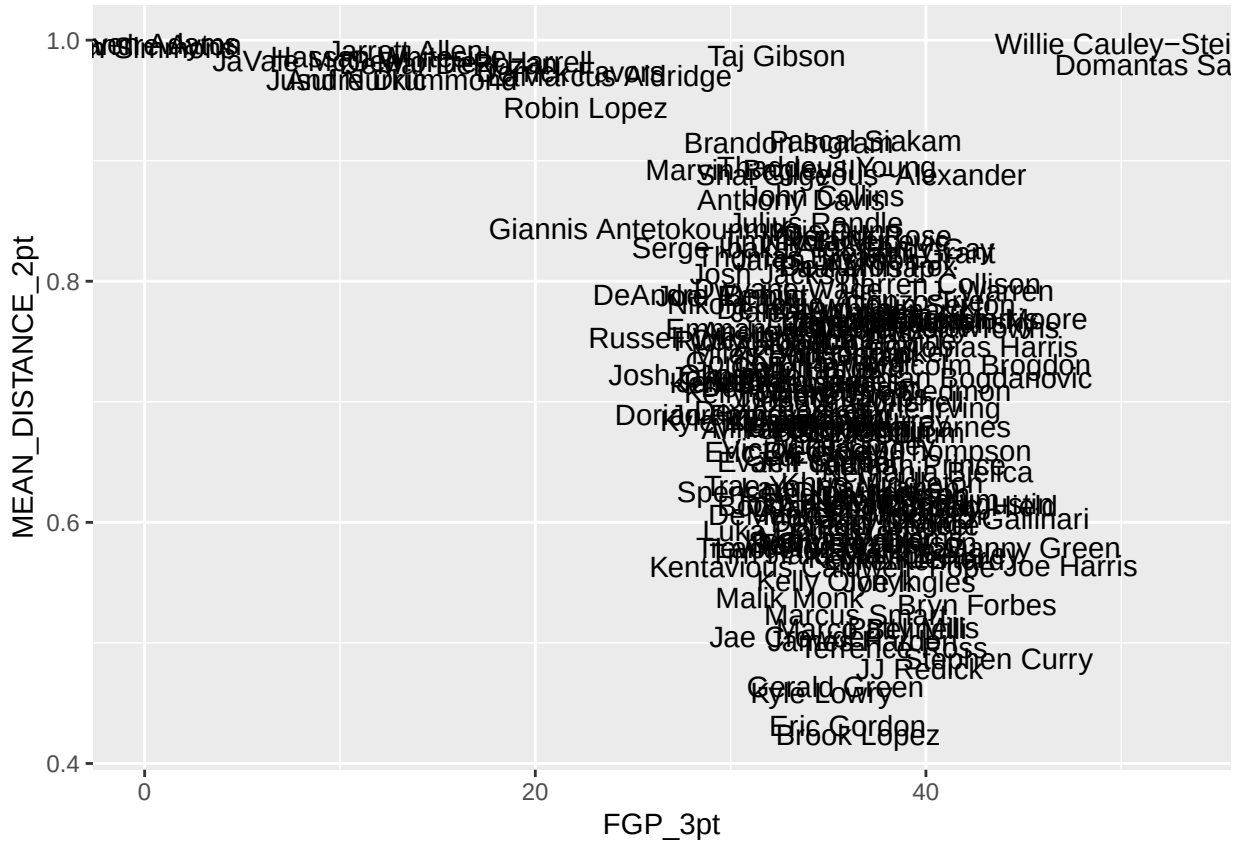
HW Q2: Create a scatterplot of these two new variables for all players with >500 total shot attempts, with 3-pt percentage on the x-axis and 2-pt distance on the y-axis. Instead of points, plot player names.

Give the plot a title and clear axis labels. (2 pts)

ScatterPlot of Players 3-Point Percentage and Mean Distance of 2pt

```
players18 %>%
  filter(FGA > 500) %>%
  ggplot(aes(x = FGP_3pt, y = MEAN_DISTANCE_2pt)) +
  geom_text(aes(label = PLAYER_NAME))
```

```
## Warning: Removed 2 rows containing missing values (`geom_text()`).
```

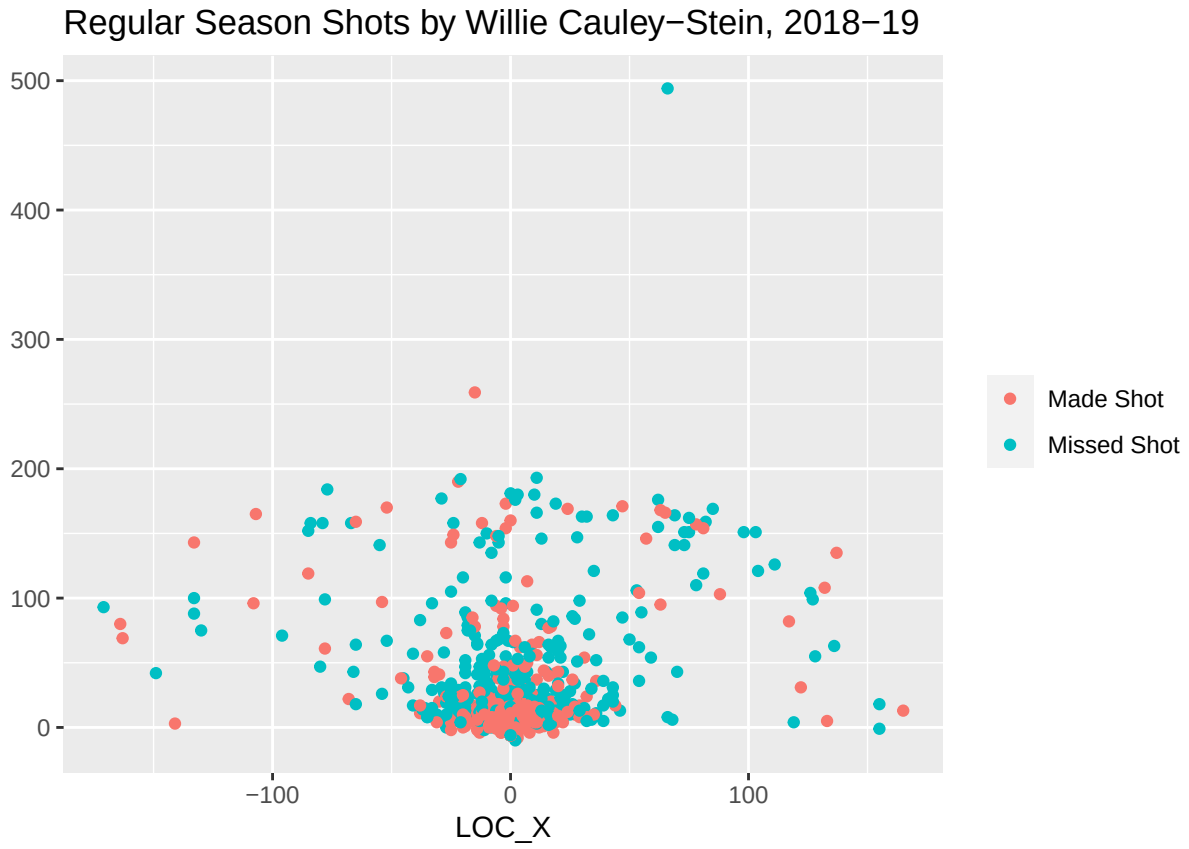


HW Q3: To whom would you award the title of Mr. Analytics for the 2018-19 NBA Season? Justify your answer in 1-3 sentences. (1 pt)

The title of Mr. Analytics would go to Willie Cauley-Stein. This would be the case because he has the highest 3pt percentage as well having the highest mean distance for 2pt. HW Q4: Create a shot chart for your awardee. Is this plot consistent with what you thought you knew about the player above? And what is his favorite type of shot(s)? Answer in 1-3 sentences. (2 pts)

Shots By Willie Cauley-Stein

```
shots18 %>%
  filter(PLAYER_NAME == "Willie Cauley-Stein") %>%
  ggplot(aes(x = LOC_X, y = LOC_Y, color = EVENT_TYPE)) +
  geom_point() +
  labs(X = "", y = "", title = "Regular Season Shots by Willie Cauley-Stein, 2018-19") +
  theme(legend.title = element_blank())
```



This is not consistent with what I thought the players shotting habits were. I thought since both values were high, he would be more of a perimeter shooter, but from the shot chart it looks like his favorite shot comes within the key. The reason his 3pt percentage is so high is that he is efficent with the 3pt shots that he takes. NOTE: In the document you submit to me, don't *show* the long code you have to use to set up the court plot. To do this, simply use the `include = FALSE` option in the chunk with that code. DO, however, show the other code you use to create the shot chart.

HW Q5: Obviously Mr. Analytics isn't a real award, and it's likely the player you identified wasn't even

in discussion for something like Most Valuable Player (MVP). Why not? Give at least 2 reasons why a Mr. Analytics (not necessarily the one you identified) might not actually generate *value* for his team? (1 pt)

The first reason that Mr. Analytics might not generate value for his team is that it doesn't show the time that he was in the game. Meaning that alot of these shots could of been when the starter of the opp team aren't in either from winning/losing by so much points. Another reason why is that these shots could of came from "garbage time" where the game is already decided so it becomes more of a shot around. **HW Q6**: Using `shots18` (NOT the data grouped by player above), calculate the overall league average 2-pt and 3-pt FG% for 2018-19. Print your results. No interpretation is needed for this question.

The League AVG in 2pt and 3pt Percentages

```
league_AVG <- shots18 %>%
  mutate(LOC_X = LOC_X/10,
         LOC_Y = LOC_Y/10,
         SHOT_DISTANCE = sqrt(LOC_X^2 + LOC_Y^2)) %>%
  mutate(pts = case_when(SHOT_TYPE == "3PT Field Goal" ~ 3,
                         TRUE ~ 2)) %>%
  group_by(TEAM_NAME) %>%
  summarize(FGA = n(),
            FGM = sum(SHOT_MADE_FLAG),
            FGA_2pt = sum(pts[pts==2])/2,
            FGA_3pt = sum(pts[pts==3])/3,
            FGM_2pt = sum(pts[pts==2]*SHOT_MADE_FLAG[pts==2])/2,
            FGM_3pt = sum(pts[pts==3]*SHOT_MADE_FLAG[pts==3])/3,
            FGP_2pt = ((FGM_2pt/FGA_2pt) * 100),
            FGP_3pt = ((FGM_3pt/FGA_3pt) * 100))

fg_2pt_avg <- mean(league_AVG$FGP_2pt)
fg_3pt_avg <- mean(league_AVG$FGP_3pt)


print(fg_2pt_avg)
```

```
## [1] 52.01981
```

```
print(fg_3pt_avg)
```

```
## [1] 35.54752
```

HINT: The `SHOT_MADE_FLAG` variable may be useful here (.5 pt).

HW Q7: Using `players18`, and among players with >50 each of 2-pt and 3-pt shot attempts, print the 6 players with the *greatest* and *smallest* differences between their FG% for 2- and 3-pt shots.

Looking at the Top 6 Players with the Greatest and Smallest differences in 2 and 3pt Shots

```
Players_Diff <- players18 %>%
  filter(FGA_3pt > 50, FGA_2pt > 50) %>%
  mutate(FG_Diff = FGP_2pt - FGP_3pt) %>%
  select(PLAYER_NAME, FGA_2pt, FGA_3pt, FGP_2pt, FGP_3pt, FG_Diff)

print(Players_Diff %>%
        arrange(FG_Diff) %>%
        head(6))
```

```
## # A tibble: 6 x 6
##   PLAYER_NAME     FGA_2pt FGA_3pt FGP_2pt FGP_3pt FG_Diff
##   <chr>             <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1 Landry Shamet        59     151    32.2    45.0 -12.8
## 2 Jevon Carter         83     102    26.5    33.3  -6.83
## 3 Allen Crabbe        114     259    34.2    37.8  -3.63
## 4 Mike Scott           71     119    38.0    41.2  -3.15
## 5 Otto Porter Jr.     121      80    47.9    48.8  -0.816
## 6 Seth Curry          214     251    46.3    45.0   1.24
```

```
print(Players_Diff %>%
        arrange(-FG_Diff) %>%
        head(6))
```

```
## # A tibble: 6 x 6
##   PLAYER_NAME           FGA_2pt FGA_3pt FGP_2pt FGP_3pt FG_Diff
##   <chr>                   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1 Dragan Bender              96     101    68.8    21.8    47.0
## 2 Derrick Favors            541      78    64.0    21.8    42.2
## 3 Dwight Powell             359     127    69.9    30.7    39.2
## 4 Giannis Antetokounmpo    1044     203    64.1    25.6    38.5
## 5 Thomas Bryant             403      99    68.5    33.3    35.2
## 6 Jonah Bolden               66      96    69.7    35.4    34.3
```

Print their names, number of shots of each type, field goal percentages of each type, and the difference between them.

Describe what you see in about 2-3 sentences. Does anyone shoot better from 3-pt range than 2-pt range? When there's a particularly large difference, is this driven by the player being way above or below average on 3-pointers, 2-pointers, both, or a combination? (1.5 pts)

Most of the people on this list are not what you would call household names besides Giannis and Seth Curry. Landry Shamet has taken alsmost 3x more shots from 3pt range than the 2pt range. This is shown with the FG_Diff being so high (negatively). This is driven by the fact also that Shamet has a higher shooting percentage from the 3pt line than within the arch.
NOTE: You may need to create (hint: mutate) several new variables for this question!

HINT: Your minimum and maximum values for the difference in shot percentages (2pt minus 3 pt) should be -0.1283 and 0.4697.

General organization and clarity of the report you turn in is worth 1 pt.