

Data Visualization HW : NBA Shooting Data

Pramote Akkapattanakul

2023-07-03

Introduction

Using NBA Shooting Data Playoff 2021 to answer below question

1. Relationship between range and success shooting rate
2. Who is the best defender among 4 players
3. Who is the best shooter among 4 players
4. Breakdown best shooter shooting chart
5. How can we improve the best shooter

Credit

For free Dataset <https://www.datacamp.com/workspace/datasets/dataset-r-nba-shooting-data>

Data Dictionary

variable	class	description
SHOOTER	character	Name of the player taking the shot
X	numeric	Horizontal distance of the shot taken from the basket in ft
Y	numeric	Vertical distance of the shot taken from the basket in ft
RANGE	numeric	Radius range of the shot taken from the basket in ft
DEFENDER	character	Name of the player defending the shot
SCORE	character	'MADE' if shot is scored, else 'MISSED'

1. Relationship between range and success shooting rate

```
library(scales)
library(jpeg)
library(tidyverse)

## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.2      v readr      2.1.4
## v forcats    1.0.0      v stringr   1.5.0
## v ggplot2    3.4.2      v tibble    3.2.1
## v lubridate  1.9.2      v tidyr     1.3.0
## v purrr      1.0.1
## -- Conflicts ----- tidyverse_conflicts() --
## x readr::col_factor() masks scales::col_factor()
## x purrr::discard()     masks scales::discard()
```

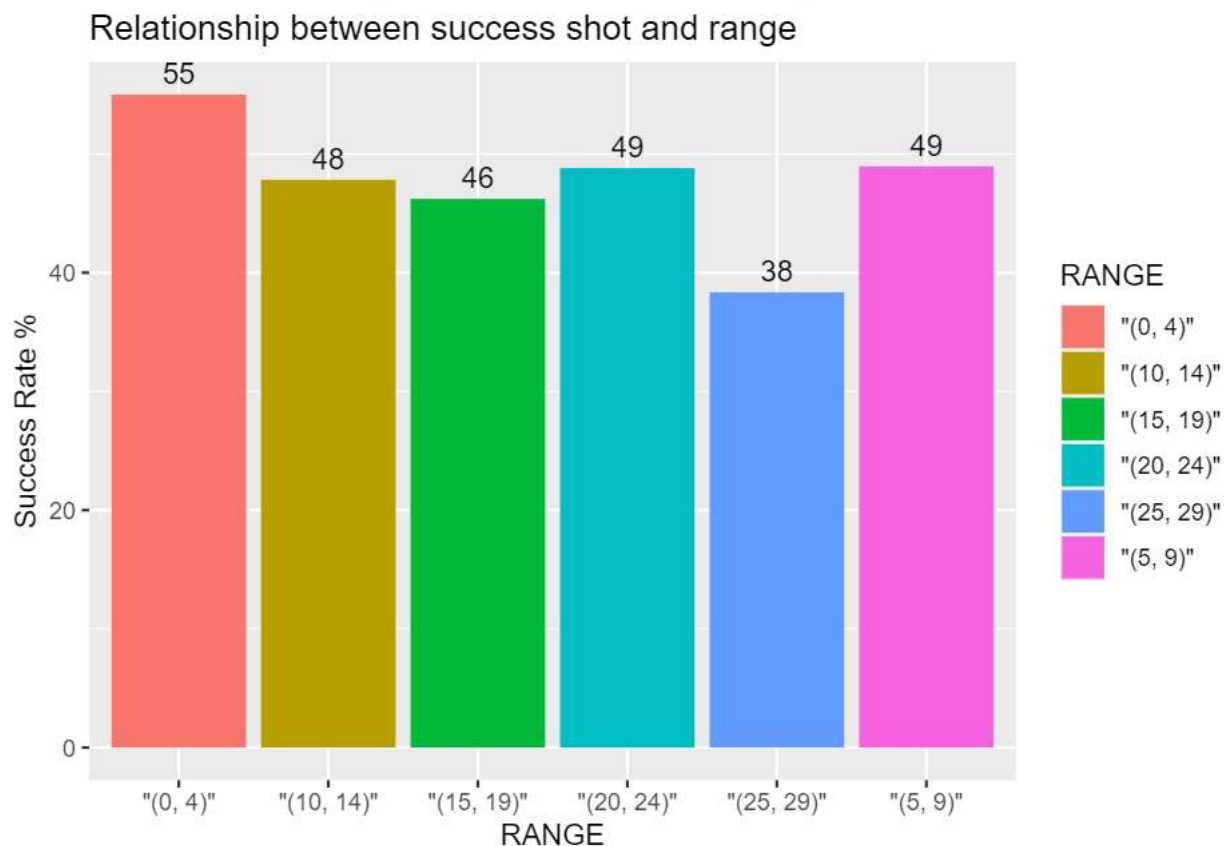
```
## x dplyr::filter()      masks stats::filter()
## x dplyr::lag()         masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(patchwork)
nba <- read_csv("datacamp_workspace_export_2023-07-03_19_52_41.csv")

## Rows: 776 Columns: 7
## -- Column specification -----
## Delimiter: ","
## chr (4): SHOOTER, RANGE, DEFENDER, SCORE
## dbl (3): INDEX, X, Y
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

range_shot <- nba %>%
  group_by(RANGE) %>%
  summarise(total_shot = n(),
            success_shot = sum(SCORE=="MADE"),
            shot_perc = (success_shot/total_shot)*100)

ggplot(range_shot, aes(RANGE, shot_perc, fill=RANGE)) +
  geom_col() +
  geom_text(aes(label = round(shot_perc)), vjust = -0.5) +
  ylab("Success Rate %") +
  labs(title = "Relationship between success shot and range")
```



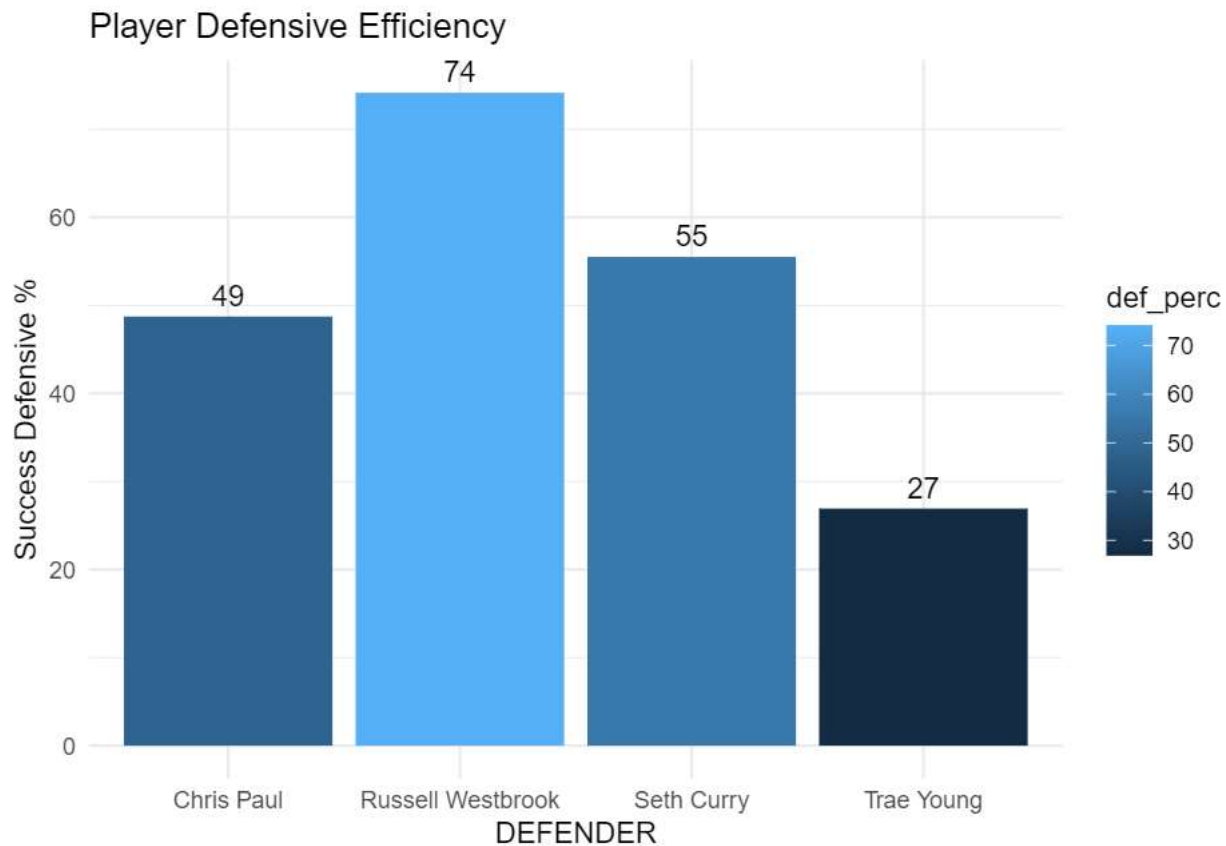
Conclusion

The shooting success rate is likely to lower when the further shooter was away from basket when shooting the ball.

2. Who is the best defender among 4 players

```
def <- nba %>%
  group_by(DEFENDER) %>%
  summarise(total_shot = n(),
            success_defense = sum(SCORE=="MISSED"),
            def_perc = (success_defense/total_shot)*100)

ggplot(def,aes(DEFENDER,def_perc,fill = def_perc))+
  geom_col()+
  geom_text(aes(label = round(def_perc)), vjust = -0.5) +
  theme_minimal()+
  ylab("Success Defensive %") +
  labs(title = "Player Defensive Efficiency")
```



Conclusion

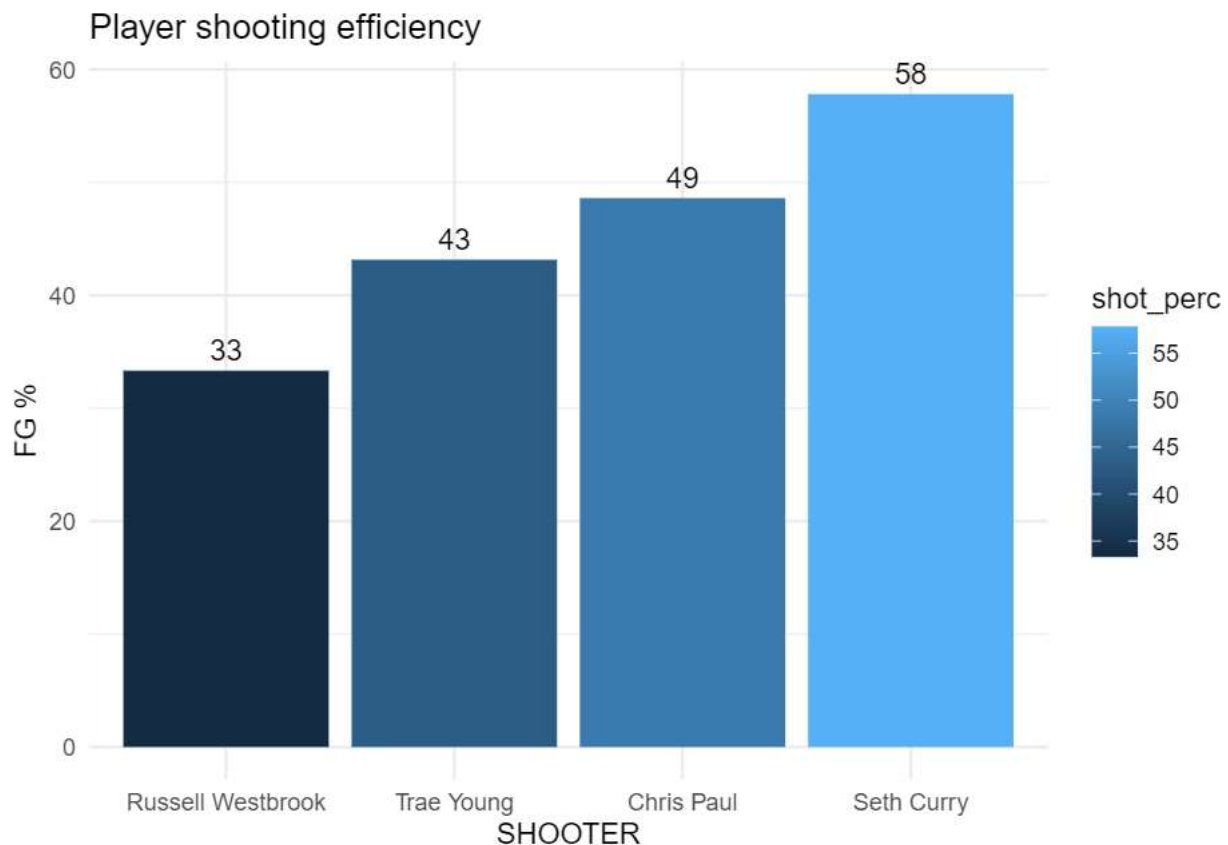
From the chart, Russell Westbrook is the best defensive player with 74% success rate on defense against each player.

3. Who is the best shooter among 4 players

```
shot_eff <- nba %>%
  group_by(SHOOTER) %>%
  summarise(total_shot = n(),
            success_shot = sum(SCORE=="MADE"),
            shot_perc = (success_shot/total_shot)*100)

shot_eff$SHOOTER<-reorder(shot_eff$SHOOTER,shot_eff$shot_perc)

ggplot(shot_eff,aes(SHOOTER,shot_perc,fill = shot_perc))+
  geom_col()+
  geom_text(aes(label = round(shot_perc)), vjust = -0.5) +
  theme_minimal()+
  ylab("FG %") +
  labs(title = "Player shooting efficiency")
```



Conclusion

Seth Curry is the best shooter with 58% field goal rate.

4. Breakdown Seth Curry shooting chart

```
court <- readJPEG("nba_court.jpg")
seth_curry <- nba %>%
  filter(SHOOTER=="Seth Curry")
```

```

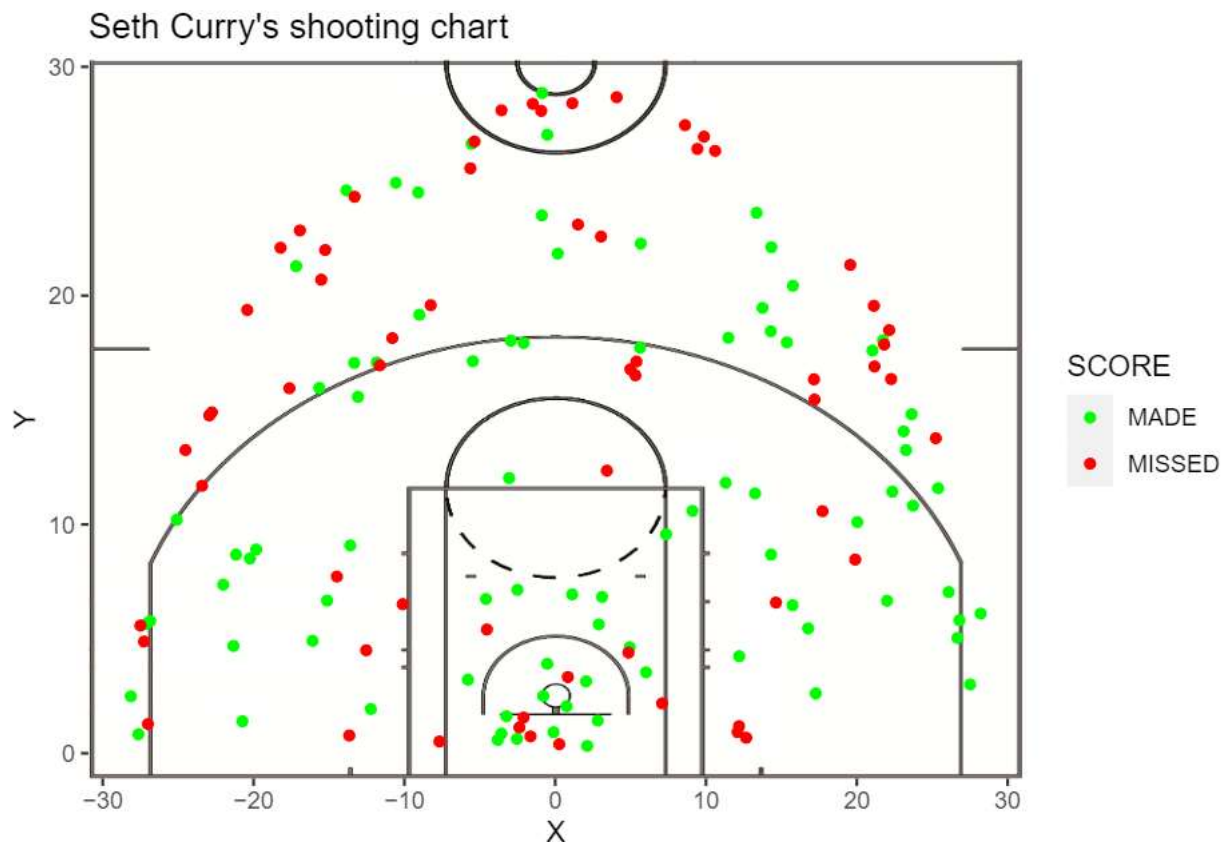
seth_curry_range <- seth_curry %>%
  group_by(RANGE) %>%
  summarise(total_shot = n(),
            success_shot = sum(SCORE=="MADE"),
            shot_perc = (success_shot/total_shot)*100)

shot_chart <- ggplot(seth_curry,aes(X,Y,color=SCORE)) +
  annotation_raster(court, xmin = -Inf, xmax = Inf, ymin = -Inf, ymax = Inf)+
  geom_point() +
  scale_color_manual(
    values = c(MADE = "green",
              MISSED = "red")) +
  labs(title = "Seth Curry's shooting chart")

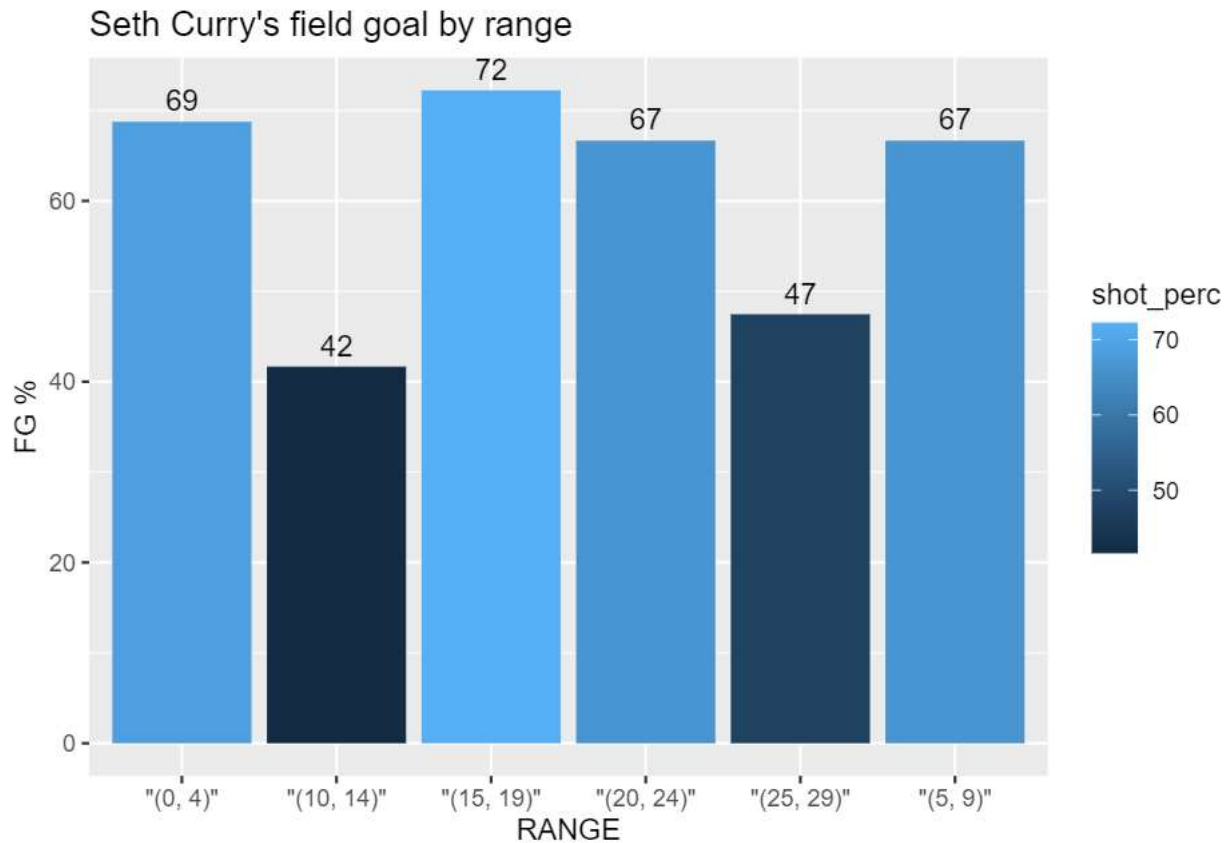
shot_range <- ggplot(seth_curry_range,aes(RANGE,shot_perc,fill=shot_perc)) +
  geom_col() +
  geom_text(aes(label = round(shot_perc)), vjust = -0.5) +
  labs(title = "Seth Curry's field goal by range")+
  ylab("FG %")

```

shot_chart



shot_range



Conclusion

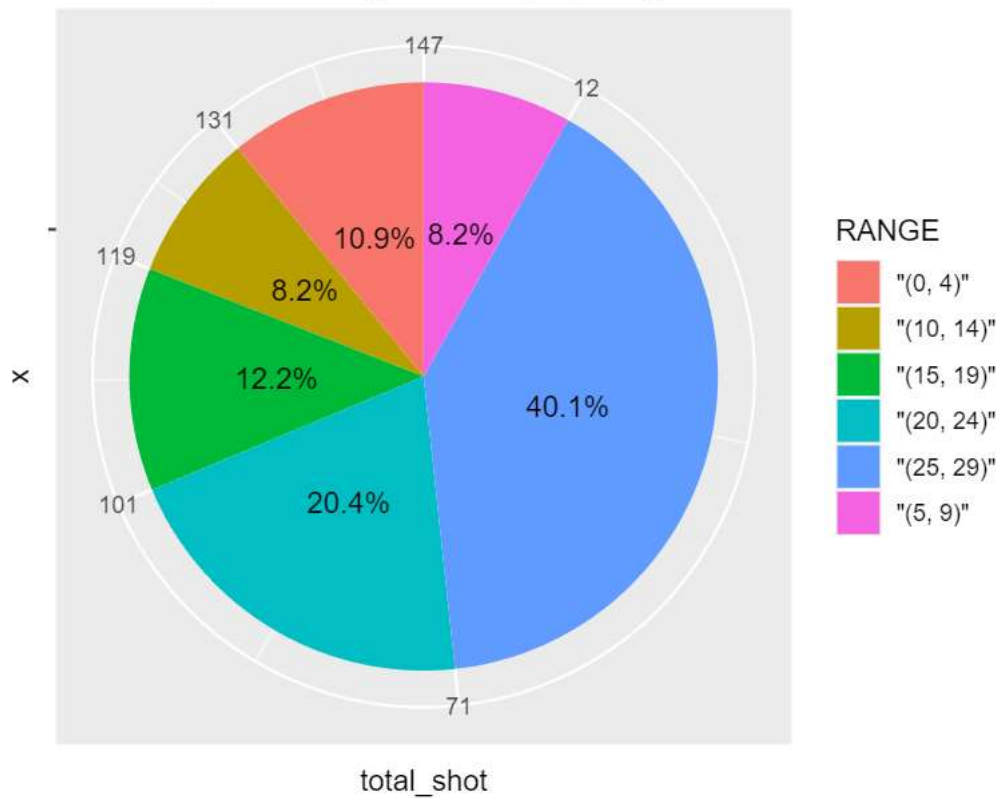
From Seth Curry's shooting chart and field goal by range, he is a great shooter at almost every range of the court.

5. How can we make his shooting efficiency even better

```
pie_chart <- ggplot(seth_curry_range,aes(x="",y=total_shot,fill = RANGE))+
  geom_bar(width = 1, stat = "identity") +
  scale_y_continuous(breaks = round(cumsum(rev(seth_curry_range$total_shot)), 1)) +
  coord_polar("y", start = 0) +
  labs(title = "Seth Curry's shooting summary by range")+
  geom_text(aes(
    label = percent(total_shot / sum(total_shot))
  ), size = 4,position = position_stack(vjust = 0.5))

pie_chart
```

Seth Curry's shooting summary by range



Conclusion

From pie chart, we can see 40.1% of total_shot made by Seth Curry is at range (25,29) with 47% field goal. It would be better for him if he moves closer to range (15,19) or (20,24) which may help to increase his field goal as currently he is doing good at both range with 72% and 67% field goal in order.