

Does three point shooting determine NCAA Tournament game outcomes more than any other factor?

Jon Michael Stroh

2021-04-20

Data Exploration (For March Madness 2021)

More3s Frequency of the winning team out shooting the losing team on 3 point shots (variable `more3s` equals 0 if losing team made more three point shots, 0.5 if an equal amount were made, and 1 if the winning team made more three point shots):

```
marchmadness2021 <- marchmadness2021 %>%  
  mutate(more3s = factor(if_else(W3P > L3P, 1, if_else(W3P == L3P, 0.5, 0))))  
  
marchmadness2021 %>% count(more3s)  
  
## # A tibble: 3 x 2  
##   more3s      n  
## * <fct>   <int>  
## 1 0         15  
## 2 0.5        9  
## 3 1         42
```

In the 2021 tournament, the winning team made more three pointers in 63.6% (42/66) of games, the teams tied in 13.6% (9/66) of games, and the losing team made more three pointers in 22.7% (15/66) of games.

Better3s Frequency of the winning team out shooting the losing team on 3 point shot percentage (variable `better3s` equals 0 if the losing team made a higher percentage of three point shots, 0.5 if an equal percentage were made, and 1 if the winning team made a higher percentage):

```
marchmadness2021 <- marchmadness2021 %>%  
  mutate(better3s = factor(if_else( (W3P/W3PA) > (L3P/L3PA),  
    1, if_else((W3P/W3PA) == (L3P/L3PA), 0.5, 0))))  
  
marchmadness2021 %>% count(better3s)
```

```
## # A tibble: 2 x 2  
##   better3s      n  
## * <fct>   <int>  
## 1 0         15  
## 2 1         51
```

In the 2021 tournament, the winning team made a higher percentage of three pointers in 77.3% (51/66) of games, while the losing team made a higher percentage in only 22.7% (15/66) of games.

```
marchmadness2021 %>% count(more3s, better3s, sort = TRUE)
```

More and Better 3s?

```
## # A tibble: 6 x 3
##   more3s better3s     n
##   <fct>  <fct>    <int>
## 1 1      1        38
## 2 0      0         9
## 3 0.5    1         7
## 4 0      1         6
## 5 1      0         4
## 6 0.5    0         2
```

In 2021, in only 13.6% (9/66) of games did the winning team make less 3s and a lower percentage of them. In contrast, in 68.2% (45/66) of games did the winning team either make more 3s and shoot a higher percentage (38 games) or tie on the number of makes and shoot a higher percentage (7 games).

More 2s Frequency of winning team out shooting losing team on 2 point shots (variable more2s equals 0 if losing team made more two point shots, 0.5 if equal amount were made, and 1 if the winning team made more two point shots):

```
marchmadness2021 <- marchmadness2021 %>%
  mutate(more2s = factor(if_else(W2P > L2P , 1, if_else(W2P == L2P, 0.5, 0))))
marchmadness2021 %>% count(more2s)
```

```
## # A tibble: 3 x 2
##   more2s     n
## * <fct> <int>
## 1 0        20
## 2 0.5       4
## 3 1        42
```

In only 42 games did the winning team make more 2 point baskets, and in 20 games the losing team made more (4 ties).

Better 2s ?? Frequency of winning team out shooting losing team on 2 point shot percentage (variable better2s equals 0 if losing team made a higher percentage of two point shots, 0.5 if equal amount were made, and 1 if the winning team a higher percentage of two point shots):

```
marchmadness2021 <- marchmadness2021 %>%
  mutate(better2s = factor(if_else((W2P/W2PA) > (L2P/L2PA), 1, if_else((W2P/W2PA) == (L2P/L2PA), 0.5, 0))))
marchmadness2021 %>% count(better2s)
```

```
## # A tibble: 3 x 2
##   better2s     n
## * <fct>    <int>
## 1 0         21
## 2 0.5        1
## 3 1         44
```

WOW- in only 44 games did the winning team out shoot the losing team on 2 point percentage, and in 21 games, the losing team out shot the winning team on 2 point percentage.

Finally FGs? morefgs Frequency of the winning team out shooting the losing team in general (more shots made) (variable morefgs equals 0 if losing team made more shots, 0.5 if equal amount were made, and 1 if the winning team made more shots):

```
marchmadness2021 <- marchmadness2021 %>%
  mutate(morefgs = factor(if_else(WFG > LFG , 1, if_else(WFG == LFG, 0.5, 0))))

marchmadness2021 %>% count(morefgs)

## # A tibble: 3 x 2
##   morefgs      n
## * <fct>    <int>
## 1 0          7
## 2 0.5        3
## 3 1         56
```

In 84.8% of games the winning team made more field goals (makes sense, only other option to win then is by making more 3s as a portion of those field goals or more free throws or both). There were 3 ties.

FG percentage? betterfgs The frequency of the winning team out shooting the losing team by percentage (higher percentage of total shots made) (variable betterfgs equals 0 if losing team made more shots, 0.5 if equal amount were made, and 1 if the winning team made more shots):

```
marchmadness2021 <- marchmadness2021 %>%
  mutate(betterfgs = factor(if_else((WFG/WFGA) > (LFG/LFGA),
    1, if_else((WFG/WFGA) == (LFG/LFGA), 0.5, 0))))

marchmadness2021 %>% count(betterfgs)

## # A tibble: 3 x 2
##   betterfgs      n
## * <fct>    <int>
## 1 0         18
## 2 0.5        1
## 3 1         47
```

However, interesting enough, in only 71.2% of games (47/66) did the winning team shoot a higher percentage on all field goals. With this being the case, clearly other factors can influence a win, the likely ones for investigation is free throw shooting (although there tends to be a small difference in this between teams in a single game), three point shooting (our variable of primary interests), and factors that can get a team more shots (offensive rebounding and forcing turnovers while limiting those things for themselves).

Data Exploration (For March Madness 2021, Additional Data)

Distribution of Shooting:

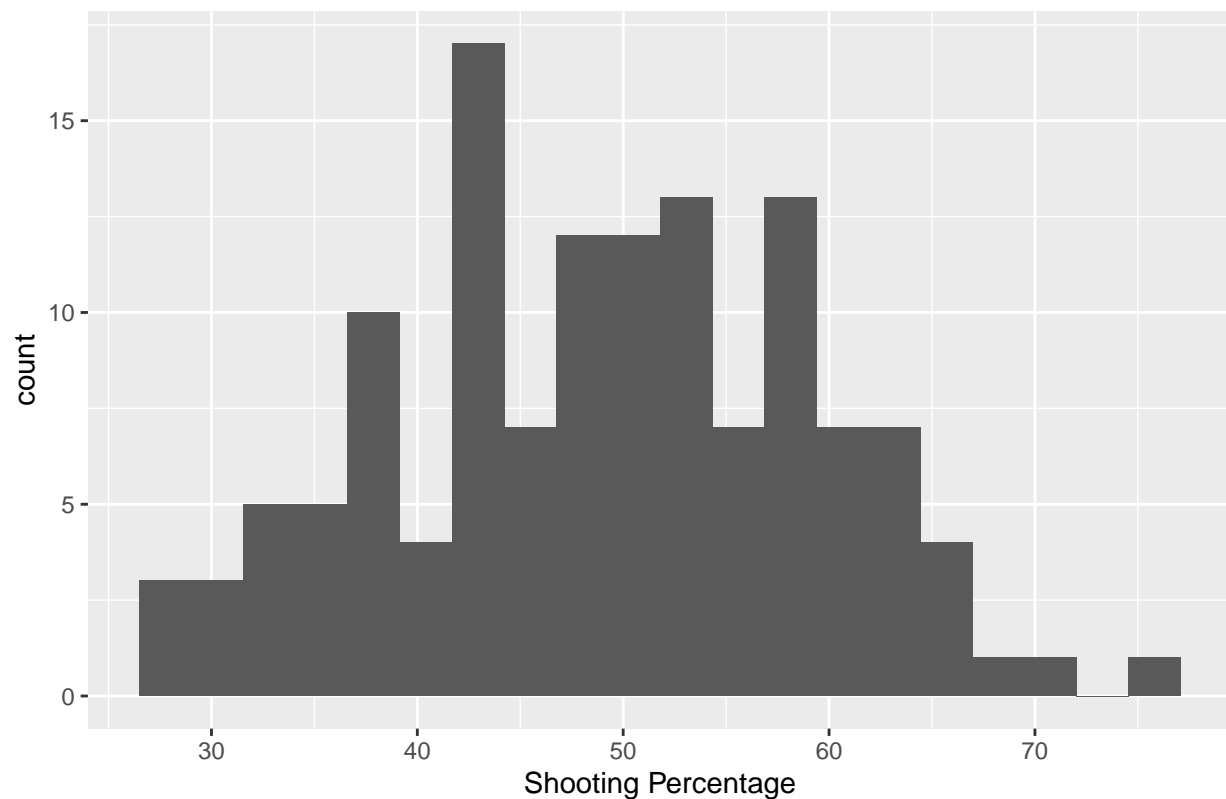
Two point shooting->

twoPercent is the two point shooting percentage by team as a percentage, rather than a decimal, for graphing.

```
mm2021byteams <- mm2021byteams %>% mutate(twoPercent = `2P%` * 100)

ggplot(data = mm2021byteams, aes(x = twoPercent)) +
  geom_histogram(bins = 20) +
  labs(x = "Shooting Percentage",
    title = "Two Point Shooting Percentages in MM 2021")
```

Two Point Shooting Percentages in MM 2021



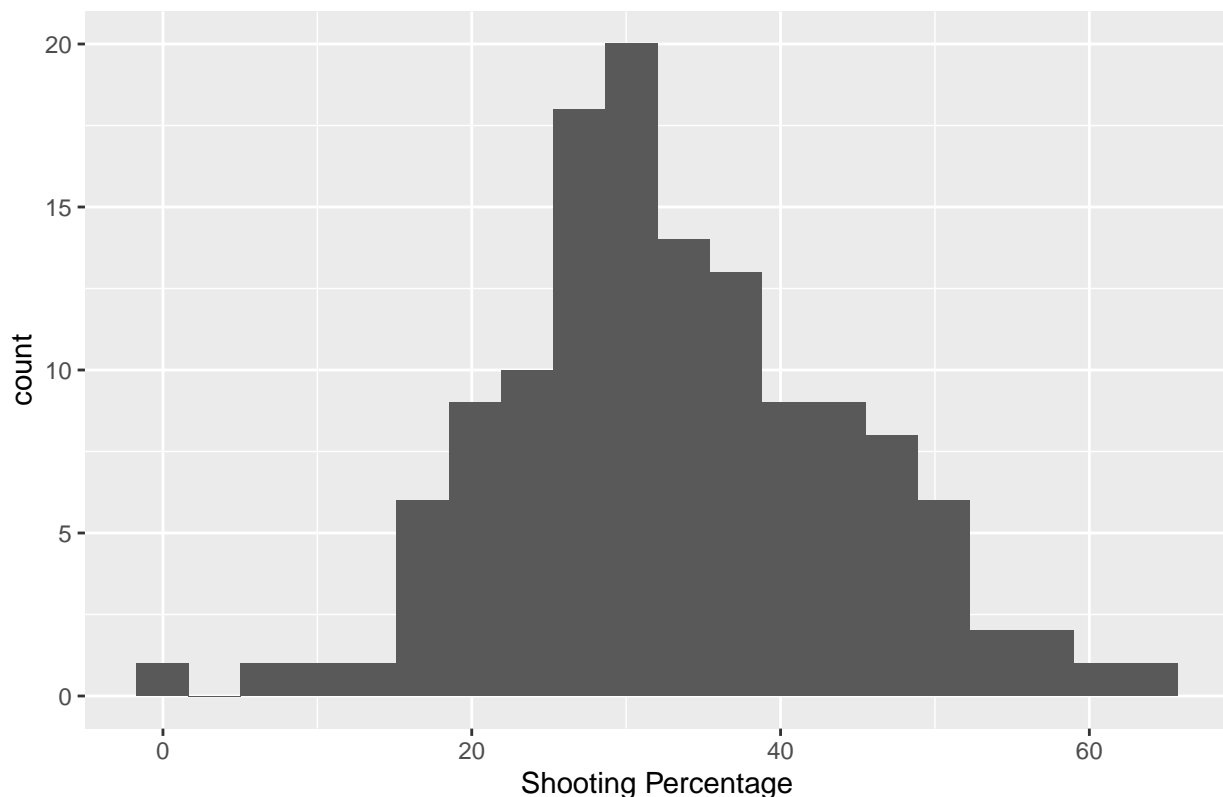
Three point shooting->

threePercent is the three point shooting percentage by team as a percentage, rather than a decimal, for graphing.

```
mm2021byteams <- mm2021byteams %>% mutate(threePercent = `3P%` * 100)

ggplot(data = mm2021byteams, aes(x = threePercent)) +
  geom_histogram(bins = 20) +
  labs(x = "Shooting Percentage",
       title = "Three Point Shooting Percentages in MM 2021")
```

Three Point Shooting Percentages in MM 2021



Comparing Shooting ->

Note: twoPercent_mean is not the shooting percent for all two point shots in the march madness tournament. It is the mean shooting two point shooting percentage for each team each game. Overall, it can not represent the percentage for total two point shooting percent because it is not adjusted for attempts by game. The same thing goes for threePercent_mean

```
mm2021byteams %>%
  summarise(twoPercent_mean = mean(twoPercent),
            threePercent_mean = mean(threePercent),
            twoPercent_sd = sd(twoPercent),
            threePercent_sd = sd(threePercent))
```

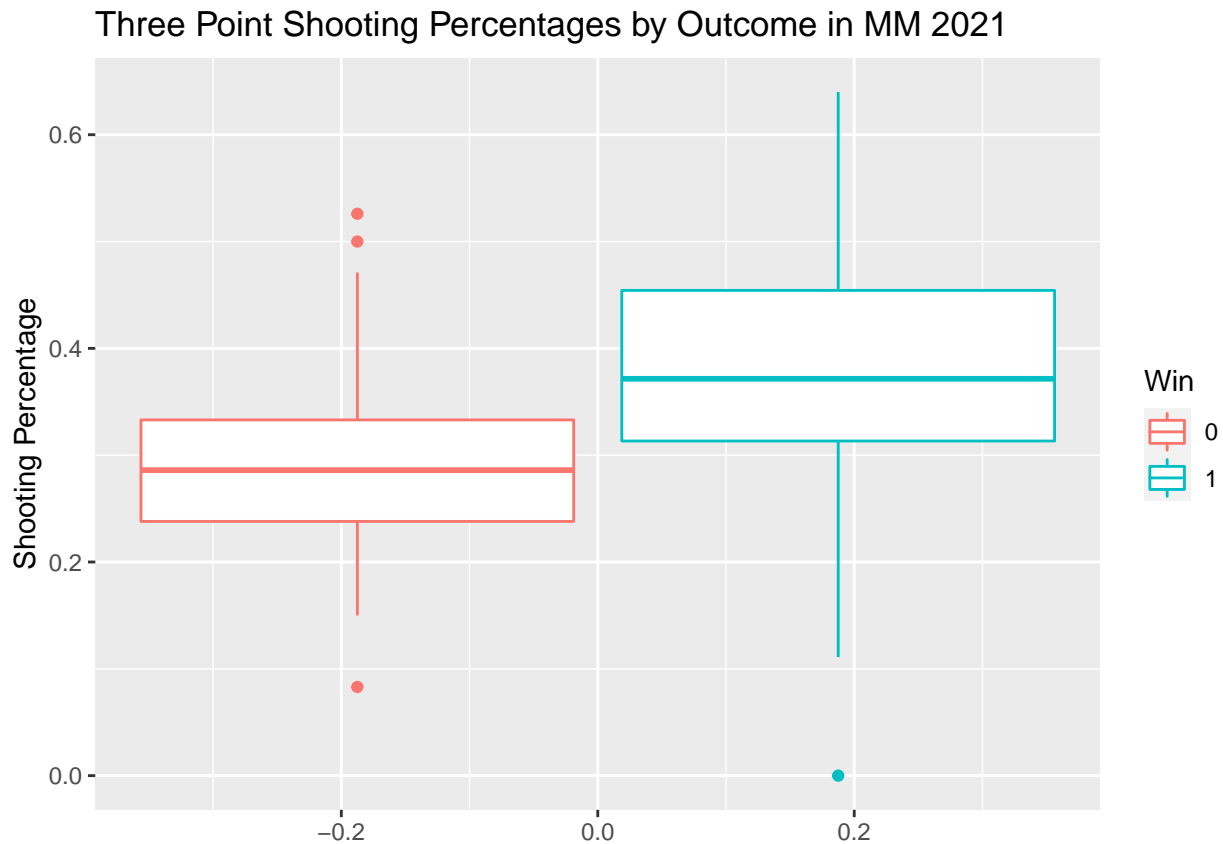
```
## # A tibble: 1 x 4
##   twoPercent_mean threePercent_mean twoPercent_sd threePercent_sd
##   <dbl>           <dbl>           <dbl>           <dbl>
## 1      49.0         33.4           10.1           11.0
```

Although three point shooting appears to much more closely resemble the normal distribution, it has a larger standard deviation indicating that there is more variation in three point shooting than two point shooting (although the standard deviations are very close) by game. According to the standard deviations, 68% of the team's three point shooting by game occurs within 10.97% of 33.45%, while 68% of the team's two point shooting by game occurs within 10.14% of 48.97%. There is slightly more variation in three point shooting performances by game than two point shooting performances.

Three Point Percentage by Outcome->

```
ggplot(data = mm2021byteams, aes(y = `3P%`, color = Win)) +
  geom_boxplot() +
```

```
labs(y = "Shooting Percentage",
     title = "Three Point Shooting Percentages by Outcome in MM 2021")
```



```
with(mm2021byteams, aggregate(~3P% ~ Win, FUN = median))
```

```
##   Win   3P%
## 1   0 0.2860
## 2   1 0.3715
```

The median three point percentage for winning teams is 37.2% compared to only 28.6% for losing teams.

Based on data exploration for all 66 games in the 2021 NCAA tournament, there seems to be a strong association between a team's three point shooting percentage for a single game and the probability that they win that game. Thus, we intend to fit explore models with three point percentage as a predictor variable for whether a team wins a game or not.