Basketball Project

Elijah Shepherd

2024-12-04

Loading the seasons

2019-2020

```
## Warning: package 'readr' was built under R version 4.2.3
```

```
## Warning: package 'dplyr' was built under R version 4.2.3
```

Creating our own columns

```
## # A tibble: 6 × 25
##   oppponent           `home/away` date    `score - a&t` `score - opp` `win/lose`
##   <chr>               <chr>       <chr>           <dbl>         <dbl> <chr>
## 1 uncg                away        11/5/2…            50            83 L
## 2 charleston southern home        11/8/2…            63            49 W
## 3 western carolina    away        11/12/…            64            90 L
## 4 utah state          away        11/15/…            54            81 L
## 5 north texas         away        11/19/…            60            80 L
## 6 nicholls st         home        11/22/…            66            54 W
## # ℹ 19 more variables: `field goals made` <dbl>, attempts <dbl>, pct <chr>,
## #   `3FG` <dbl>, `3FG - attempts` <dbl>, PCT...12 <chr>, FT <dbl>,
## #   `FT - attempts` <dbl>, PCT...15 <chr>, OFF <dbl>, DEF <dbl>, TOT <dbl>,
## #   PF <dbl>, AST <dbl>, TO <dbl>, BLK <dbl>, STL <dbl>, PTS <dbl>,
## #   High_scoring <chr>
```

Outlining the goals

Basketball, like a lot of sport competitions, is played to win. A high level basketball team like our very own A&T men’s team could possibly use a better breakdown of what is affecting their ability to win. We wanted to look at our given data set for the 2019-2020 season and try to determine which variables of ours had the biggest impact on our ability to win games.

Let’s start easy

The first thing we wanted to look at from our chart is the amount of points we were scoring per game. A very redundant but important part of the game is the ability to score a lot of points. Below we have created graphs that look at our average points per game on a density plot to see how they affect our wins and losses

```
## Warning: package 'ggplot2' was built under R version 4.2.3
```

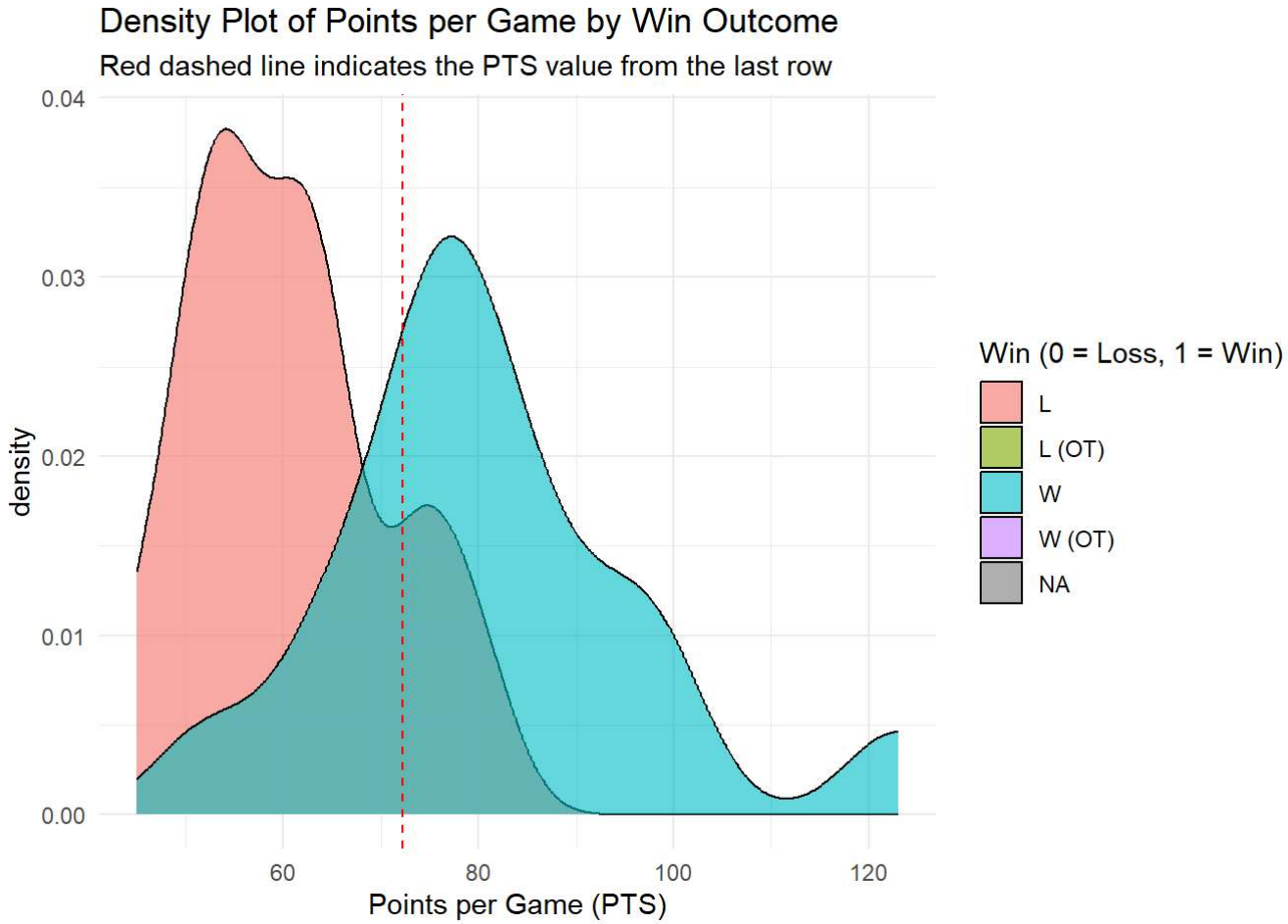
```
## Warning: Removed 1 row containing non-finite outside the scale range
## (`stat_density()`).
```

```
## Warning: Groups with fewer than two data points have been dropped.
## Groups with fewer than two data points have been dropped.
## Groups with fewer than two data points have been dropped.
```

```
## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf

## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf

## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf
```



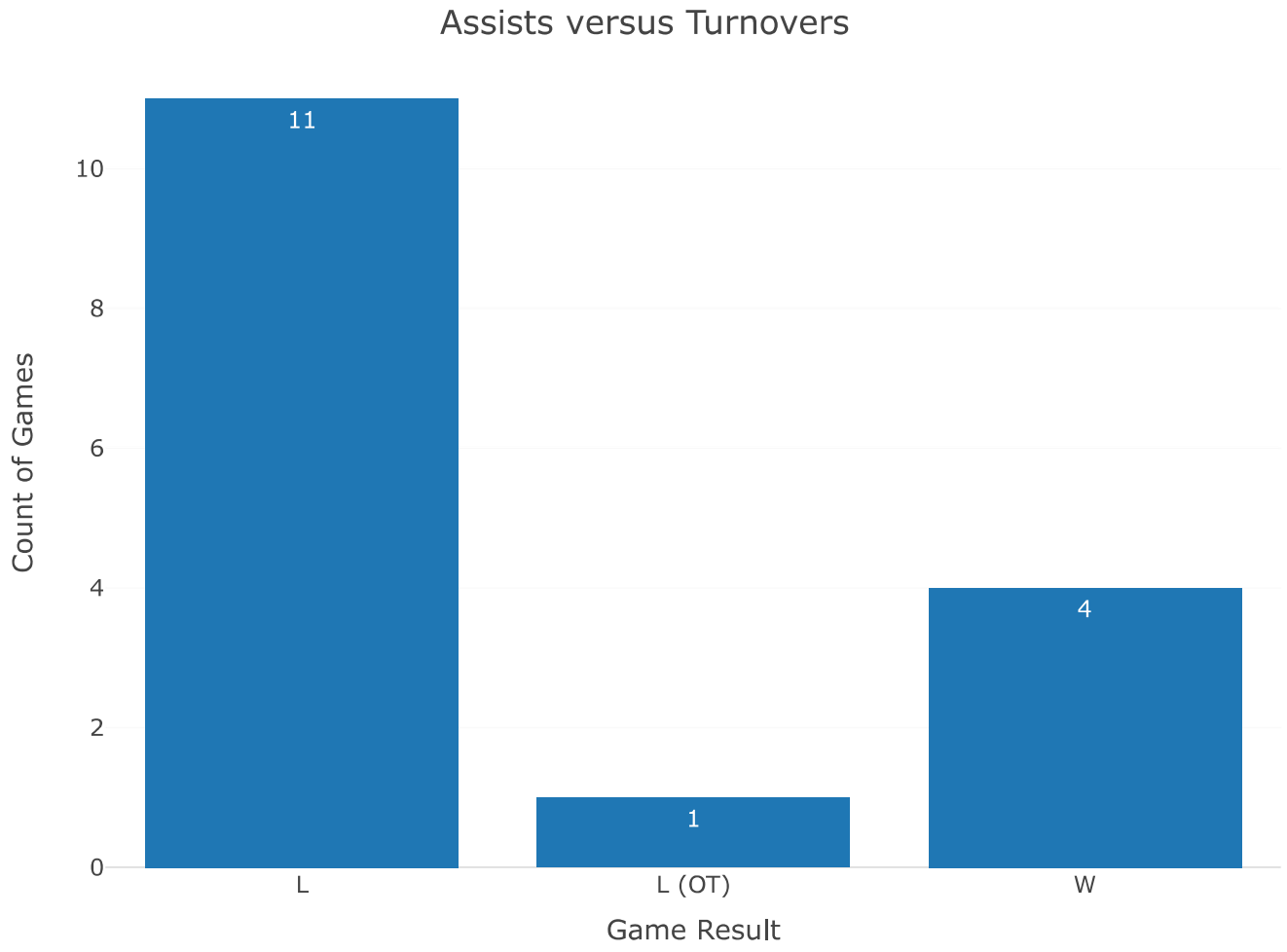
We have created a density plot to look at the distribution of the wins (blue) and losses (pink). We can see that our wins portion of the graph has a wider range than our pink portion of the graph, which tells us that the amount of points we score in wins has more variability than those in losses. Let's also notice the peak of each curve. For the losses category, the peak of the curve is at about 50 points per game. While the wins curve peak is at around 70-75. We can use this to help us say that our higher scoring games are usually more associated with winning.

Creating a graph based on whether we won or lost the game when our turnovers outnumbered our assists.

Based on our previous information we can see that We had a slight turnover problem this year. We had 15 games where our turnovers outnumbered our actual number of assists. We can even hover over the points on the graph and see what team it was against. Did this turnover problem cause more wins or losses? Below is the code to figure out how many games where our turnovers were more than our assists that we won or lost.

```
## Warning: package 'plotly' was built under R version 4.2.3
```

```
## Warning: Ignoring 1 observations
```

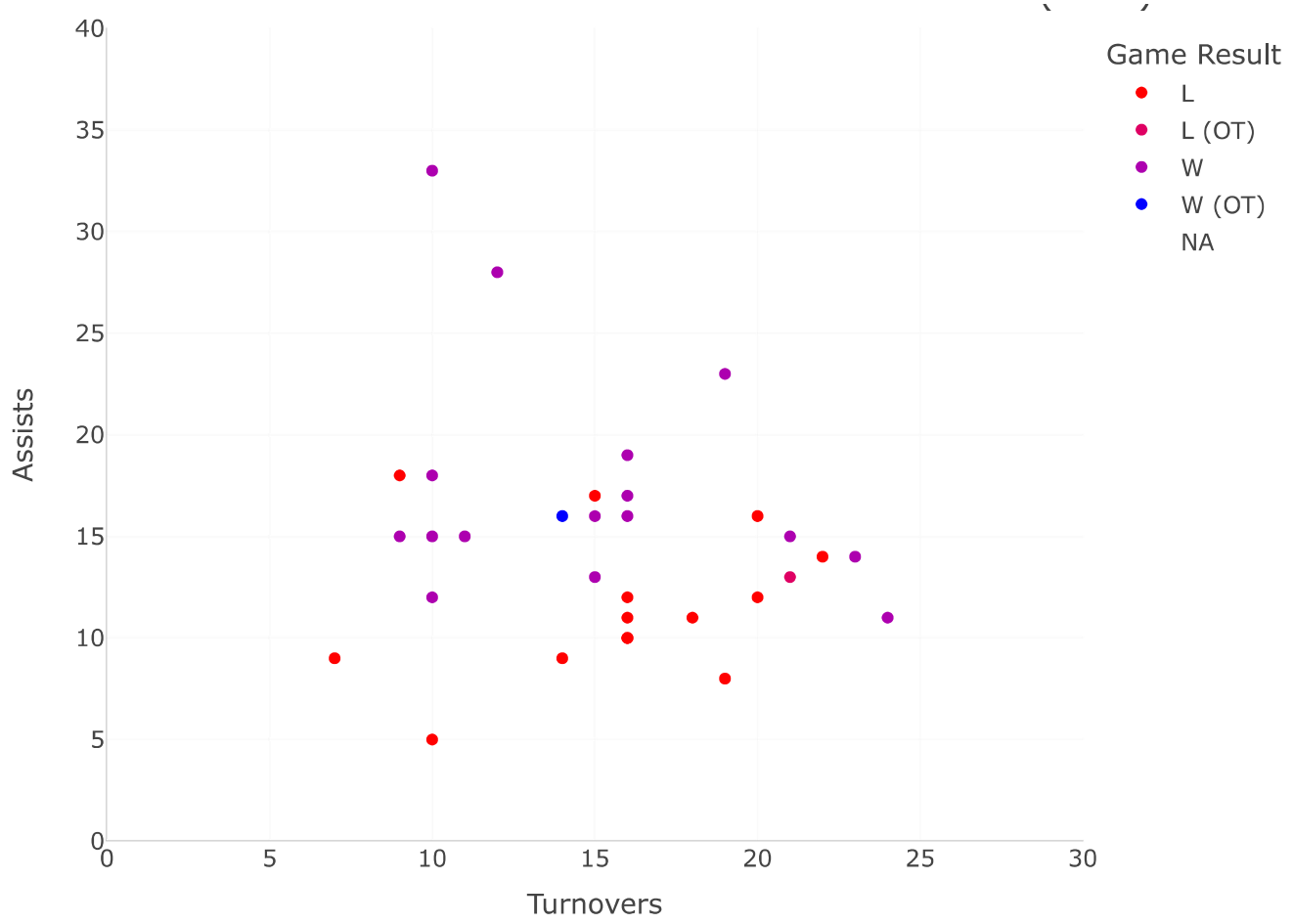


As we can see, we have 12 total games where when our Turnovers outnumber our assists, we lose. There were only 4 games in which when this happened, we actually ended up winning the game. This can be used for a potential coach who wants to know just how turnovers are affecting his team's ability to win games.

Now we want to put these together. Let's say you wanted to see what the specific games were that we lost and what was the turnover count for that game while also being able to see whether or not we won or lost.

```
## Warning: Ignoring 1 observations
```

Assists > Turnovers: Outcomes and Turnover Counts (2019)



Now we have a graph that shows us the whether or not we won or lost, the opponent we played, and our actual number of turnovers for the given game. We can see from the graph that when we have a high assist game, even if our turnovers are high, we have a better chance at when the game. Whenever we had a low assist AND high turnover game, we were more likely to lose the game.

Creating an effectve filed goal percentage column

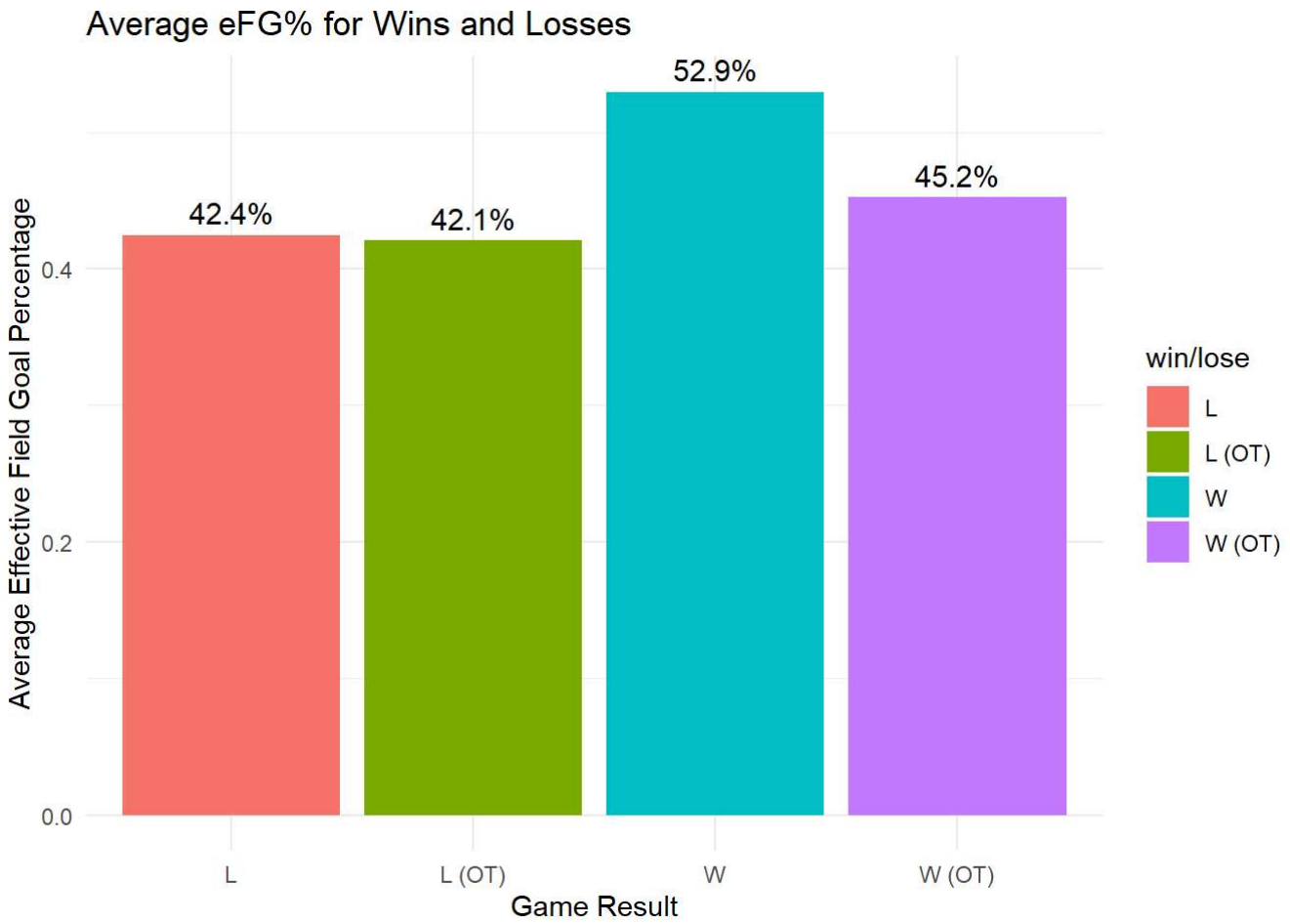
Winning is the name of the game, and in order to do that you must be able to shoot the ball well. In basketball, any attempt to make a basket is tracked as a field goal. This field goal attempt can come from anywhere on the court, as long as it is during live play (so free throws are not included). A general way to track how well a team shoots is by tracking their field goal percentage (can be calculated by $Number\ \ of\ \ makes\ \ /\ \ Number\ \ of\ \ attempts$). However, this is not always the best metric to use though, as the three point shot in basketball is worth three points, while everything else is worth two points.

So we wanted to add the Effective field goal percentage (efg) metric. It is known as an advanced basketball metric which counts for the appropriate weight of the 3-point shot and can allow a team to better truly understand how they shoot in total as a team.

```
## # A tibble: 34 × 25
##    oppponent           `home/away` date   `score - a&t` `score - opp` `win/lose`
##    <chr>               <chr>       <chr>          <dbl>         <dbl> <chr>
##  1 uncg                away        11/5/…            50            83 L
##  2 charleston southern home        11/8/…            63            49 W
##  3 western carolina    away        11/12…            64            90 L
##  4 utah state          away        11/15…            54            81 L
##  5 north texas         away        11/19…            60            80 L
##  6 nicholls st         home        11/22…            66            54 W
##  7 eastern michigan    away        11/24…            54            58 L
##  8 jacksonville        home        11/30…            45            61 L
##  9 longwood            home        12/4/…            52            41 W
## 10 bradley             away        12/7/…            52            83 L
## # ℹ 24 more rows
## # ℹ 19 more variables: `field goals made` <dbl>, attempts <dbl>, pct <chr>,
## #   `3FG` <dbl>, `3FG - attempts` <dbl>, PCT...12 <chr>, FT <dbl>,
## #   `FT - attempts` <dbl>, PCT...15 <chr>, OFF <dbl>, DEF <dbl>, TOT <dbl>,
## #   PF <dbl>, AST <dbl>, TO <dbl>, BLK <dbl>, STL <dbl>, PTS <dbl>,
## #   efg_percentage <dbl>
```

Now the the metric has been added to the original dataset, we wanted to determine how does this affect winning? As mentioned earlier, we believe that shooting well as a team is vital to being a winning basketball program. So we want to see does our thought process hold up.

Does efg percentage affect winning?

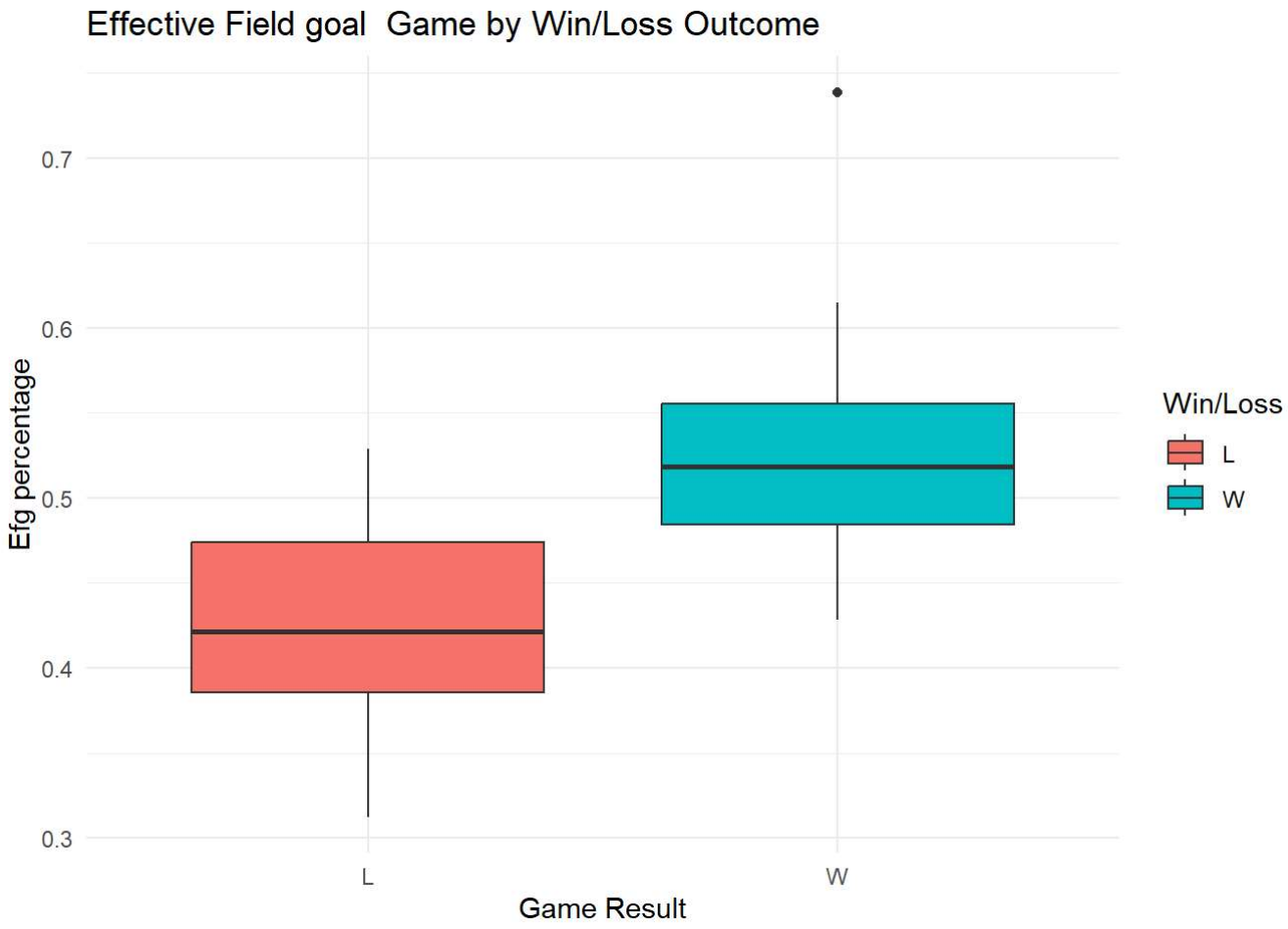


This is a helpful bar graph to showcase just how effective field goal percentage affects winning. As you can see from the code above we calculated the average efg for wins and losses and plotted them on a bar graph. You'll notice that losses generally had a lower efg than wins did, in fact the average efg for a win was just over a 10 value increase than one for a loss.

Finally, let's look at the distribution for our wins and losses based on our effective field goal percentage metric. We wanted to break this down into a boxplot so that we can get a better idea of what our distributions look like.

```
# Filter dataset to include only rows with 'W' or 'L' in the 'win/lose' column
efg_WL <- efg[efg$`win/lose` %in% c("W", "L"), ]

ggplot(efg_WL, aes(x = factor(`win/lose`), y = efg_percentage, fill = factor(`win/lose`))) +
  geom_boxplot() +
  labs(
    title = "Effective Field goal  Game by Win/Loss Outcome",
    x = "Game Result",
    y = "Efg percentage",
    fill = "Win/Loss"
  ) +
  theme_minimal()
```



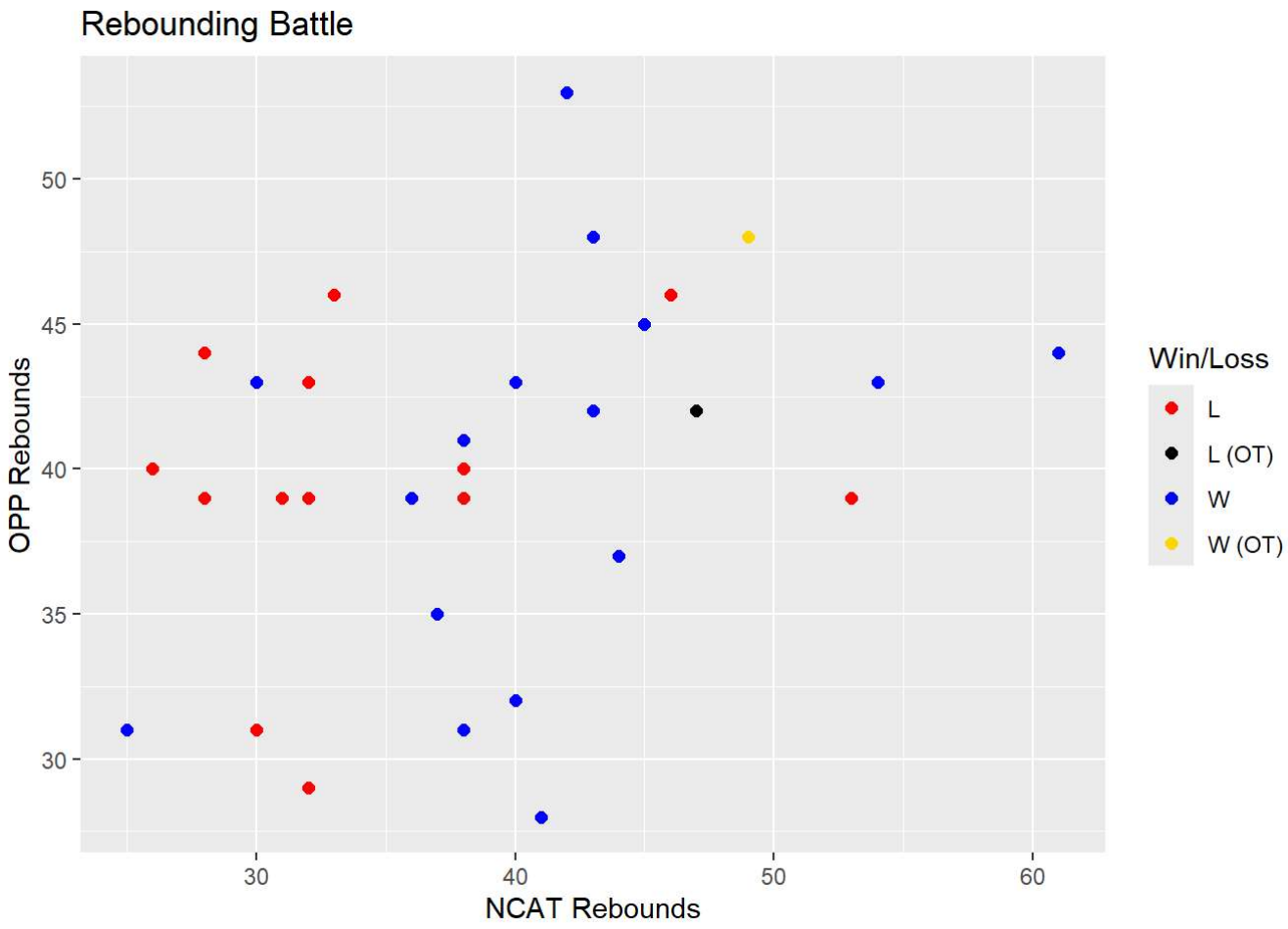
As we can see, the median for our wins is a few percentage points higher than our losses. Also, there seems to be a wider distribution of our percentages in losses than wins, however even our upper Q3 for our losses is lower than our lower quartile for our wins. Basically, when we our effective field goal percentage is low, we can conclude from this boxplopt that we are more than likely to lose.

```
## Warning: package 'tidyverse' was built under R version 4.2.3
```

```
## Warning: package 'tibble' was built under R version 4.2.3
```

```
## Warning: package 'purrr' was built under R version 4.2.3
```

```
## Warning: package 'lubridate' was built under R version 4.2.3
```


Rebounding Battle
OPP Rebounds
50
45
40
35
30
30
40
50
60
NCAT Rebounds
Win/Loss
L
L (OT)
W
W (OT)