Drive Killers and Unforced Errors

Brent Satterwhite

5/10/2020

Introduction

We often hear that in football, penalties are drive killers. I will be doing an investigation on how penalties truly effect drives and how costly they can be in certain situations.

Drive Killers

The transformations below collect each drive and summarise with some descriptive statistics. Penalties are categorized from an offensive point of view, on whether they were comitted by the team with possession AKA by the offense (`penalty_committed`) or whether they were made by the defense (`penalty_awarded`).

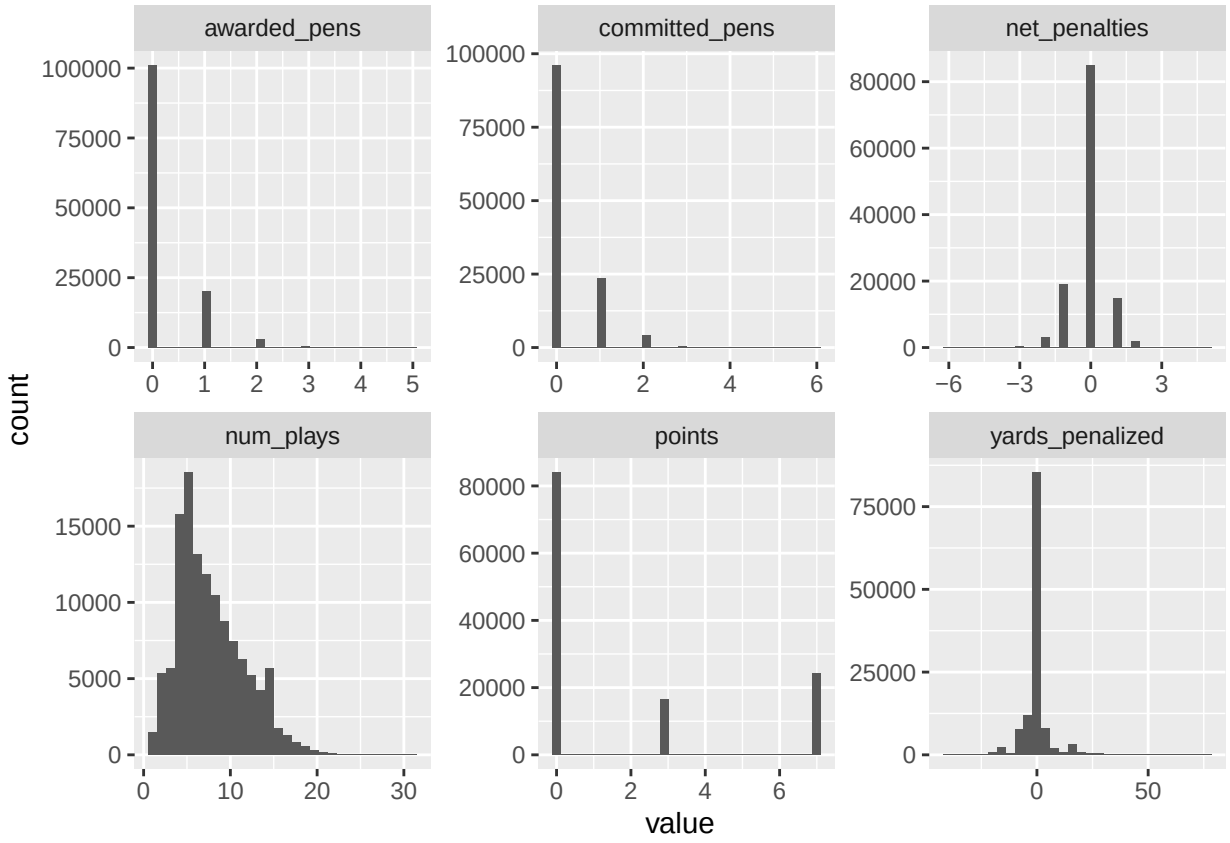
```
pbp_with_penalties <-
    pbp %>%
    mutate(
        penalty_committed = if_else(penalty & penalty_team == posteam, 1, 0),
        penalty_awarded = if_else(penalty & penalty_team != posteam, 1, 0)
    )

drives <-
    pbp_with_penalties %>%
    mutate(
        drive_score = if_else(drive_how_ended == "Touchdown", 7,
                              if_else(drive_how_ended == "Field_Goal", 3, 0))) %>%
    group_by(game_id, drive) %>%
    summarise(
        #success = max(drive_score) / 7,
        points = max(drive_score),
        committed_pens = sum(penalty_committed),
        awarded_pens = sum(penalty_awarded),
        net_penalties = awarded_pens - committed_pens,
        yards_penalized = max(drive_yards_penalized),
        num_plays = n()) %>%
    ungroup() %>%
    filter(!is.na(points)) %>%
    select(-game_id, -drive)
head(drives)
```

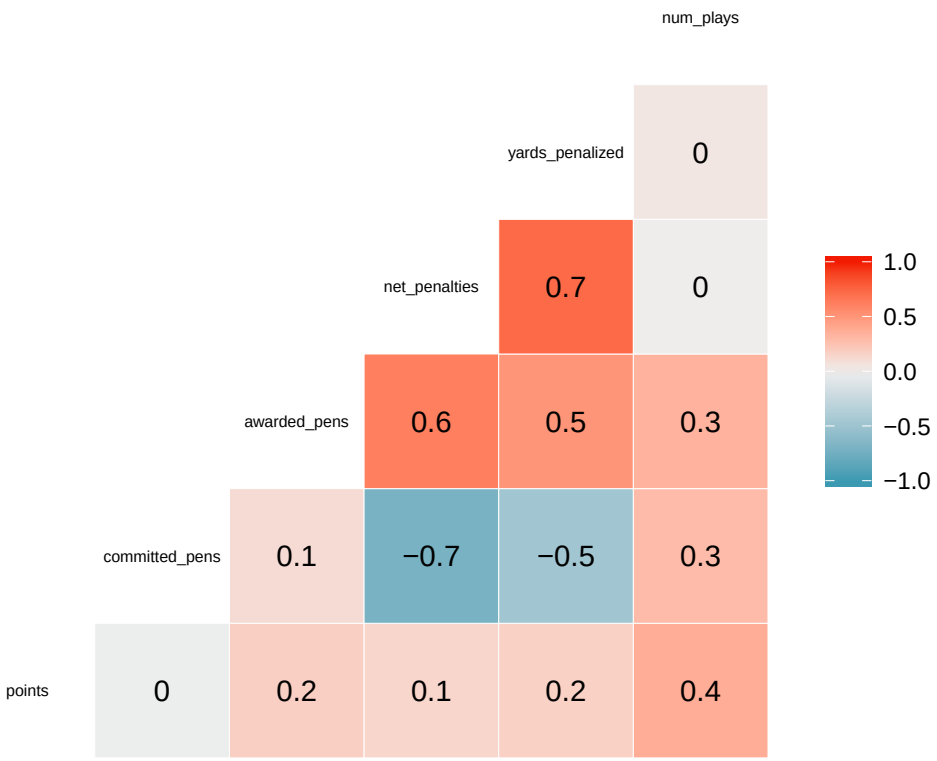
```
## # A tibble: 6 x 6
##   points committed_pens awarded_pens net_penalties yards_penalized
##    <dbl>          <dbl>        <dbl>         <dbl>           <dbl>
## 1      0              0            0             0               0
## 2      0              0            0             0               0
## 3      0              0            0             0               0
## 4      7              0            0             0              -5
## 5      0              0            0             0               0
## 6      0              0            0             0               0
## # ... with 1 more variable: num_plays <int>
```

Histograms for each column show that on a typical drive, no penalties are committed by either team:

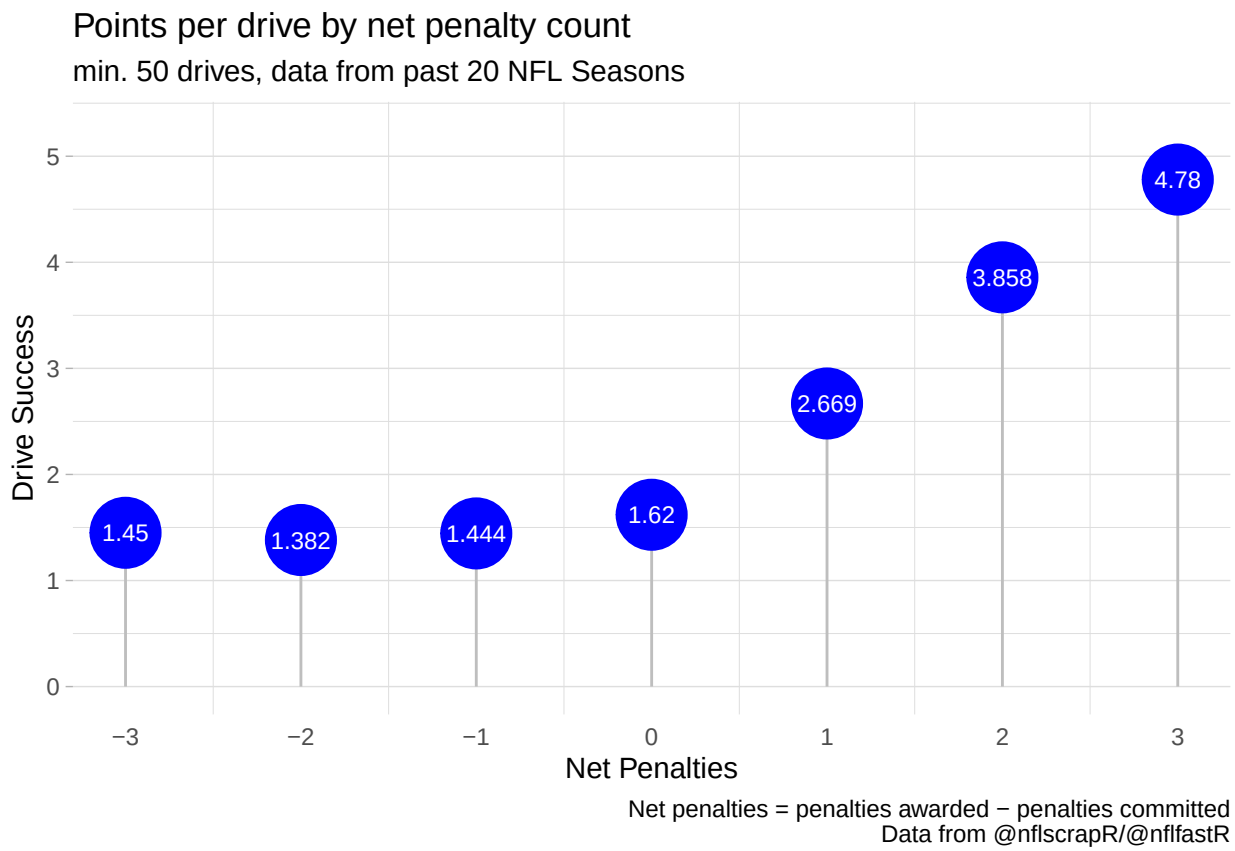
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



One concern for looking at the data at a drive level, is that the longer a drive goes on, the more likely it is that penalties will occur. So this could artificially increase the number of drives where the offense scored despite committing one or more penalties. Using `ggcorr()`, we can create a correlation grid to show how correlated the different factors are that we're dealing with.

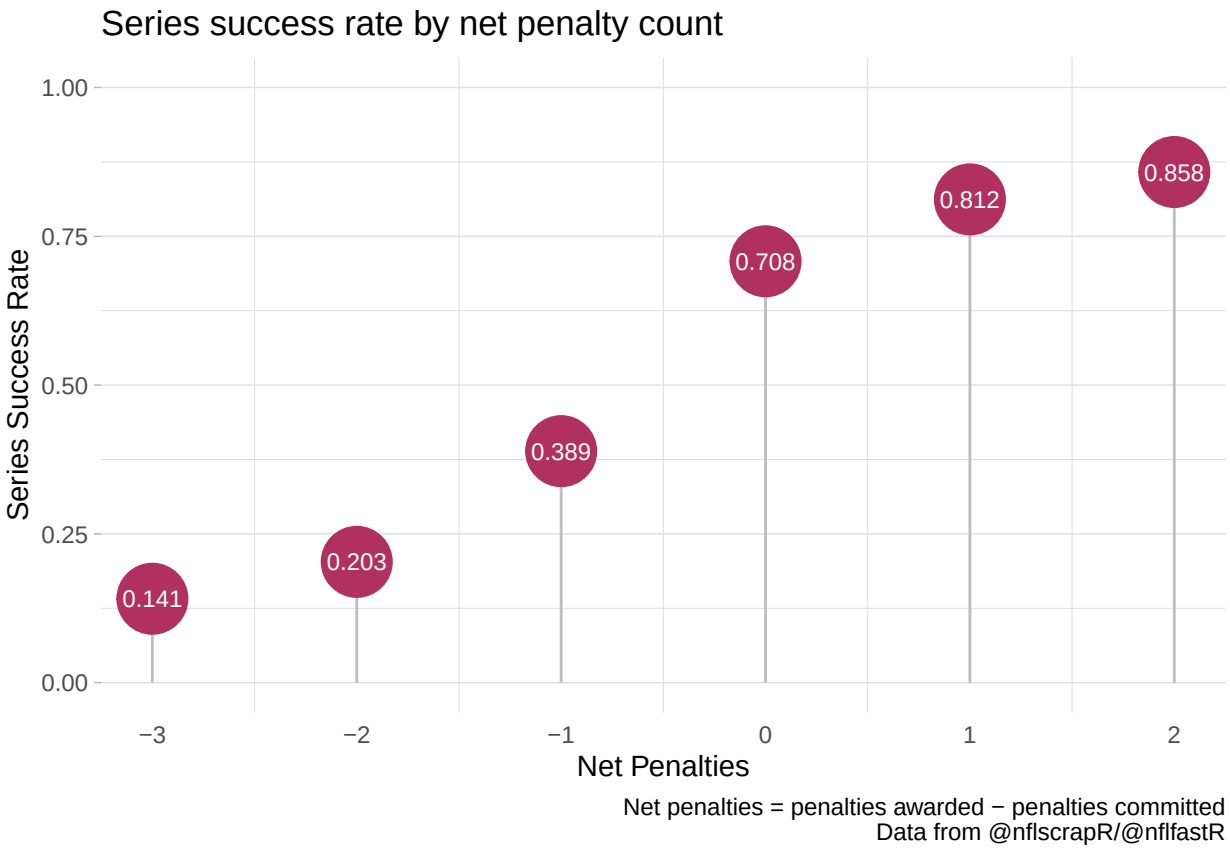


Finally, we can distill this data into points per drive for each number of net penalties.



When looking at a drive-by-drive basis, defensive penalties represent a huge advantage for the offense. Beyond the baseline at zero, every additional net penalty is an extra point per drive, while moving below zero shows the data to be pretty flat. I imagine this influenced heavily by the fact that the longer the drive, the more penalties there may be.

To mitigate this effect, I'll look at the data in a very similar way but at a series level instead of a drive level. See https://github.com/leesharpe/nfldata/blob/master/UPDATING-NFLFASTR.md#add-in-columns-for-series-data for definition of series and series success



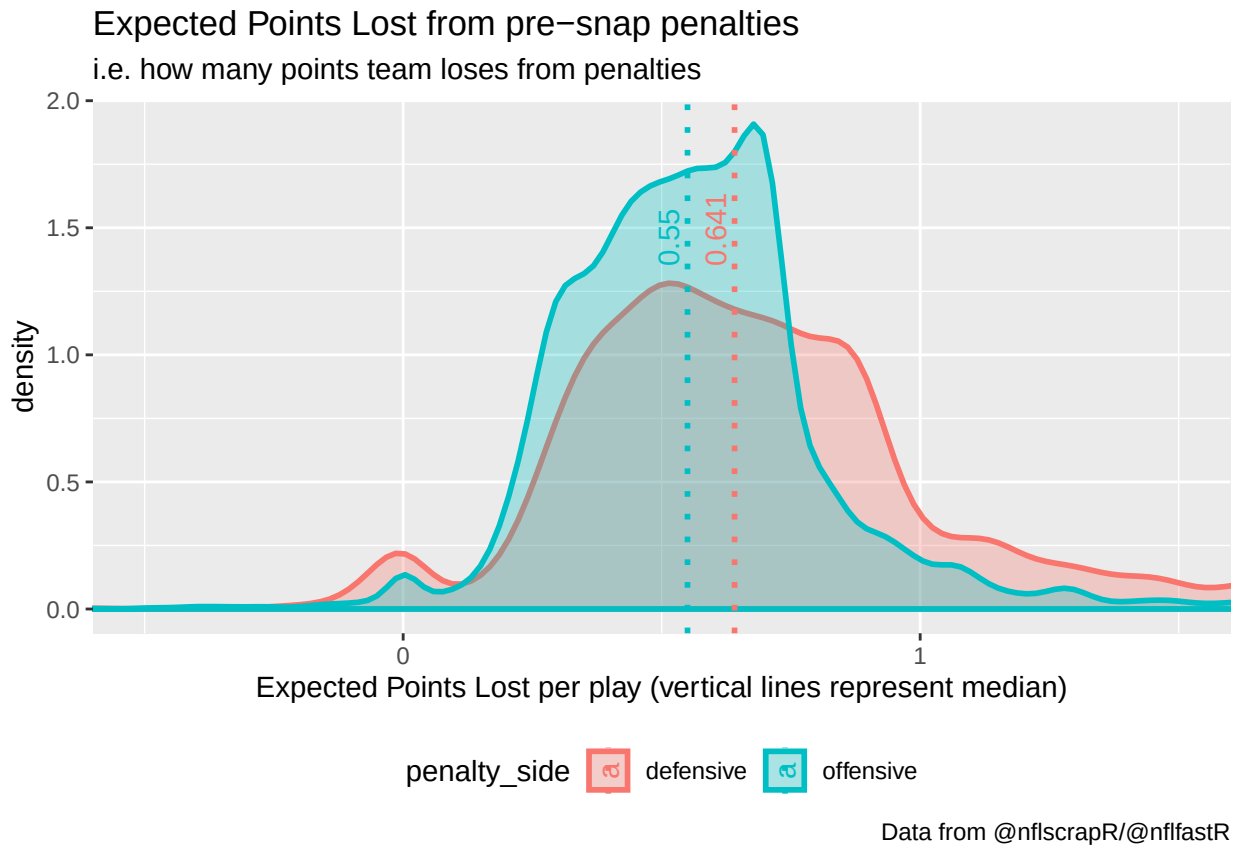
With a `0.708` success rate as the baseline, an offense clearly has an advantage in a given series. Yet if they commit just one additional penalty than the defense, that success rate plummets to `0.389`.

Unforced Errors

In this section, I want to look at how costly mistakes can be for the offense or defense. I'm filtering down to presnap penalties because those are typically self-inflicted mistakes that could have been avoided.

The table below shows the frequency of each pre-snap penalty occurring and the following chart shows how costly a pre-snap penalty is per play for offensive compared to defense.I've added a column called `epl` for expected points lost, which simply negates the EPA for offensive penalties, since EPA is built from an offensive point of view.

penalty_type	occurrences	penalty_side
False Start	12270	offensive
Defensive Offside	3585	defensive
Delay of Game	2365	offensive
Neutral Zone Infraction	1644	defensive
Encroachment	1178	defensive
Illegal Formation	878	offensive
Illegal Shift	451	offensive
Illegal Motion	261	offensive
Illegal Substitution	249	offensive
Defensive Too Many Men on Field	74	defensive
Defensive Delay of Game	58	offensive
Offensive Offside	43	offensive
Offensive Too Many Men on Field	20	offensive



On average, defensive penalties pre-snap are more costly than offensive penalties pre-snap, but this is likely mitigated due to how common false starts are.

Conclusion

At a series level, penalties can be very costly. A single offensive penalty can completely shift the advantage the offense typically has. Drives tell a slightly different story, but due to the confounding factors of penalty count, drive length, and points in a drive, I trust the series data more.

For future research, I want to look at how penalties have affected individual teams and which teams have lost or gained the most points from penalty